

**Dissertation**

submitted to the  
Combined Faculty for the Natural Sciences and Mathematics  
of the Heidelberg University, Germany

for the degree of  
Doctor of Natural Sciences

Put forward by

**Josefine Handrack**

born in: Saarbrücken, Germany

Oral examination: 29.11.2019



# MR-based treatment planning for proton therapy

Referees: Prof. Dr. Peter Bachert  
Prof. Dr. Oliver Jäkel





## MR-based treatment planning for proton therapy

Magnetic resonance imaging for image guidance of proton therapy could greatly reduce dosimetric uncertainties during treatment, as uncertainties due to set-up or anatomical changes can be detected and corrected for. For dose calculation, however, MR images lack information on the stopping power of the tissue. The use of synthetic computed tomography (CT) images derived from MR images was thus investigated using the example of prostate cancer in this thesis.

First, the impact of bulk density (BD) overwrites in the planning CT for small ( $< 5$ ) sets of tissue classes was investigated. For the modified CT image with BD values for five tissues, sufficient dosimetric agreement (mean gamma pass rates (GPRs) above 98.5 % for the PTV) and mean absolute range difference below 1.8 mm were observed.

The same set of tissue classes was subsequently identified in the rigidly registered MR image and overwritten with the same BD values as before. The mean GPR decreased to 95.2 %, and the absolute range difference increased to 4.0 mm. The observed differences are unacceptable and emphasize the necessity of MR guidance for proton therapy of prostate cancer.

Using a deformably registered MR image to overcome the anatomical differences between the CT and the MR, the dosimetric agreement was found clinically acceptable for most cases (mean GPR = 97.6 %), and the mean absolute range difference decreased to 2.7 mm.

The workflow implemented in this thesis allows fast and robust generation of synthetic CT images for proton therapy. The presented conversion technique is compatible with different MR scanners and can be extended to different anatomical sites with little extra effort. While the full potential of MR images has not yet been tapped, synthetic CT images are a useful tool for MR-guided proton therapy.



## MR-basierte Bestrahlungsplanung in der Protonentherapie

Magnetresonanz (MR)-Bildgebung hat großes Potential, Unsicherheiten in der Protonentherapie zu reduzieren. Dazu zählen Positionierungsunsicherheiten und anatomische Veränderungen im Patienten, die durch das MR Bild detektiert und korrigiert werden könnten. Dem MR Bild fehlt jedoch eine entscheidende Information für die Dosisberechnung: das Bremsvermögen des Gewebes. Daher wird in dieser Arbeit die Erstellung und Anwendung von sogenannten synthetischen Computertomographie (CT)-Bildern am Beispiel von Prostatakrebs untersucht.

Im ersten Schritt wurde eine geringe Anzahl ( $< 5$ ) an Gewebeklassen im Planungs-CT mit entsprechenden Gruppendichte-Werten überschrieben. Das modifizierte CT-Bild mit genau fünf Gewebeklassen zeigte ausreichende dosimetrische Übereinstimmung mit dem Planungs-CT (die mittlere Quote der Voxel mit akzeptabler Abweichung gemäß Gamma-Analyse, vom Englischen "Gamma Pass Rate (GPR)", lag über 98.5 % für das PTV) und der mittlere, absolute Reichweitenunterschied lag bei maximal 1.8 mm.

Anschließend wurden die gleichen fünf Gewebeklassen aus dem rigide registrierten MR-Bild klassifiziert und mit den entsprechenden Gruppendichte-Werten überschrieben. Die mittlere GPR sank auf 95.2 % und die mittleren, absoluten Reichweitenunterschiede stiegen auf 4.0 mm. Diese Unterschiede zum Planungs-CT sind inakzeptabel und unterstreichen die Notwendigkeit für MR-geführte Protonentherapie bei Prostatakrebs.

Da große anatomische Veränderungen zwischen den CT- und den entsprechenden MR-Bildern zu beobachten waren, wurden im letzten Schritt synthetische CT Bildern aus deformierbar registrierten MR Bildern erstellt. In den meisten Fällen waren die Bestrahlungspläne dieser synthetischen CTs klinisch akzeptabel (mittlere GPR = 97.6 %), gleichzeitig sanken die mittleren, absoluten Reichweitenunterschiede auf 2.7 mm.

Der hier implementierte Ablauf ermöglicht die schnelle und robuste Erzeugung von synthetischen CT Bildern und deren Analyse im Vergleich mit dem Planungs-CT. Die präsentierte Methode ist kompatibel mit unterschiedlichen MR-Scannern und kann mit wenig Mehraufwand auf weitere Körperregionen erweitert werden. Das Potential der MR-geführten Protonentherapie ist noch nicht vollständig ausgeschöpft, jedoch sind synthetische CT-Bilder ein wichtiger Schritt und ein nützliches Werkzeug auf dem Weg dahin.



# Contents

<b>List of acronyms</b>	<b>iv</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. Physics background</b>	<b>3</b>
2.1. Magnetic Resonance . . . . .	3
2.1.1. The MR Experiment . . . . .	4
2.1.2. Artefacts in MR . . . . .	7
2.2. Computed Tomography . . . . .	7
2.3. Radiotherapy . . . . .	10
2.3.1. Intensity modulated radiotherapy . . . . .	11
2.3.2. Intensity modulated proton therapy . . . . .	15
2.3.3. Uncertainties in radiotherapy . . . . .	17
<b>3. MR-based dose calculation</b>	<b>19</b>
3.1. Motivation . . . . .	19
3.2. Available Methods . . . . .	20
3.3. Goal of this thesis . . . . .	21
<b>4. Material and Methods</b>	<b>25</b>
4.1. Patient data . . . . .	25
4.2. Treatment planning tool: matRad . . . . .	29
4.3. Rigid and Deformable Registrations . . . . .	31
4.4. Bias Field Correction . . . . .	32
4.5. Own developments for this thesis . . . . .	33
4.5.1. Bulk Density Computed Tomographies . . . . .	33
4.5.2. Determination of bulk density values . . . . .	36
4.5.3. Generation of synthetic CTs . . . . .	38
4.6. Dose calculation and re-calculation . . . . .	41
4.7. Analysis Metrics . . . . .	42
4.8. Uncertainty Sampling . . . . .	46
<b>5. Results</b>	<b>49</b>
5.1. BD-CTs for photon therapy . . . . .	49
5.1.1. Discussion . . . . .	56
5.2. BD-CTs for proton therapy . . . . .	59
5.2.1. HIT dataset . . . . .	59
5.2.2. Gold Atlas dataset . . . . .	69

5.2.3. Discussion . . . . .	76
5.3. sCT from MR . . . . .	81
5.3.1. Range Analysis . . . . .	81
5.3.2. Gamma Analysis . . . . .	82
5.3.3. Dose Volume Histogram Analysis . . . . .	83
5.3.4. Discussion . . . . .	87
<b>6. Summary and Outlook</b>	<b>93</b>
 <b>Appendices</b>	 <b>95</b>
<b>A. Deformable Image Registration</b>	<b>97</b>
<b>B. List of Publications</b>	<b>99</b>
<b>Bibliography</b>	<b>101</b>



# List of acronyms

2D	two-dimensional
3D	three-dimensional
BD	bulk density
BD-CT	bulk density computed tomography
BEV	beam's-eye-view
BFC	bias field correction
CS	Compton scattering
CT	computed tomography
CTV	clinical target volume
DD	dose distribution
DECT	dual-energy x-ray CT
dicom	Digital Imaging and Communications in Medicine
DVH	dose volume histogram
DVM	dose volume metric
FOV	field-of-view
GPR	gamma pass rate
GTV	gross tumour volume
HIT	Heidelberg Ion-Beam Therapy Center
HLUT	Hounsfield look-up table
HU	Hounsfield unit
ICRU	International Commission on Radiation Units & Measurements
IMPT	intensity-modulated proton therapy
IMRT	intensity-modulated radiotherapy
IQR	interquartile range
linac	linear accelerator
MAE	mean absolute error
MCS	multiple Coulomb scattering
MR	magnetic resonance
OAR	organ at risk
pCT	planning computed tomography
PD	proton density



---

PE	photoelectric effect
PET	positron emission tomography
PP	pair production
PTV	planning target volume
RBE	relative biological effectiveness
rD	radiological depth
rED	relative electron density
RF	radio-frequency
RPW	rectal posterior wall
sCT	synthetic computed tomography
SOBP	spread-out Bragg peak
SPR	stopping power ratio
st	soft tissue
UTE	ultra short echo time
VOI	volume of interest
WE	water-equivalent
WEPL	water-equivalent path length
ZTE	zero echo time



# 1. Introduction

Radiotherapy is one of the main therapeutic interventions in cancer therapy: estimated 50% of all cancer patients would benefit from it (Borras et al., 2015). Although radiotherapy with protons is less available compared to conventional photon therapy, the stopping power of charged particles in matter can be beneficial for treatment of immobile and localized tumours. Protons deposit most of their energy, meaning they cause most cell damage, at the end of their range (Wilson, 1946), the so-called “Bragg peak”. Therefore less beam angles are necessary to cover the target, which better spares healthy tissue. To place the Bragg peak at the correct position for therapy, knowledge of the energy absorption properties (dominated by the electron density) of the tissue in the beam path is necessary. For this reason, computed tomography (CT) images, which measure the attenuation of ionizing x-ray radiation, provide the current gold standard for estimating of the electron density with highest geometrical accuracy.

To treat less static tumours, repeated or even online imaging is desirable. This motivates the use of magnetic resonance (MR) images for treatment planning, while additionally sparing the dose burden of CT imaging. Up-to-date MR images could be used for dose recalculation from the original treatment plan to detect a possible need for action, or treatment planning itself. However, electron density cannot be assessed by means of MR, which measures the magnetization of protons in the tissue in the presence of magnetic fields. The difference between a CT and an MR image is visualized in figure 1.1. A detailed description of the two imaging modalities can be found in section 2. In proton therapy, the need for up-to-date or live information on electron density distribution is especially crucial, due to the steep dose gradients of the proton beams: changes in electron density caused by anatomical or geometrical changes in the beam path can lead to severe changes in the position of the Bragg peak, leading to underdosage of the tumour or overdosage of healthy tissue. Examples for such changes in anatomy are organ motion during the treatment, daily positioning errors, tumour shrinkage or patient weight loss in the course of the treatment (Lomax, 2008). While MR imaging is superior in detecting these changes, the electron density as the basis for dose calculation, however, has to be included in the image by other means.

One solution for including electron density information into an MR image is the generation of so-called “synthetic computed tomography (sCT)” images. SCTs have been investigated in many studies during the last years (comprehensive reviews have

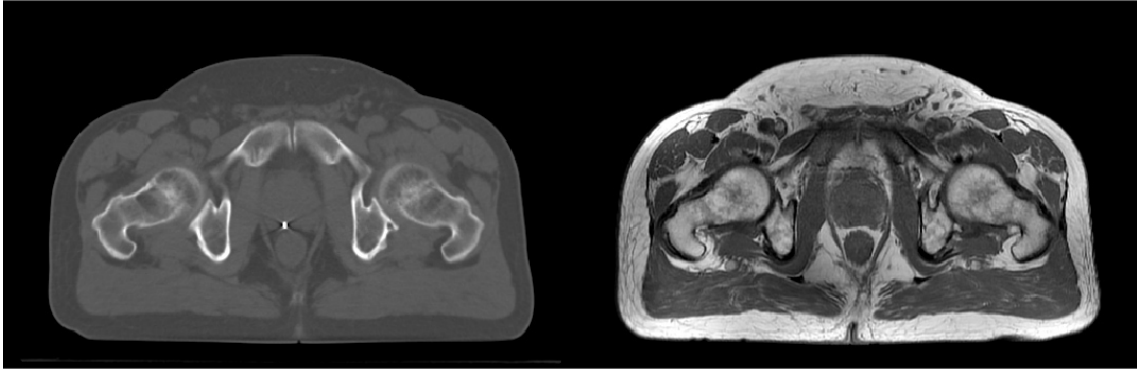


Figure 1.1.: Left: Axial slice of a CT scan from a pelvic cancer patient. Right: Corresponding slice to the CT slice of a T2 weighted MR scan.

been published by Johnstone et al. (2017) and Edmund and Nyholm (2017)), mostly for photon therapy applications. For conventional radiotherapy of pelvic cancer, commercial products (Tyagi et al., 2017; Siversson et al., 2015) are available on the market, as well as integrated into combined MR imaging and radiotherapy devices (so-called "MR-linacs"). For proton therapy, less literature is available. Approaches with different levels of complexity have been applied: from simple bulk density (BD) overwrite techniques (Maspero et al., 2017), over the dual model by Koivula et al. (2016), to more complex machine learning techniques (Rank et al., 2013; Edmund et al., 2014; Pileggi et al., 2018; Spadea et al., 2019). More details on the available techniques can be found in section 3.2.

The aim of this thesis was the generation of sCT images for proton therapy of prostate cancer. Prostate cancer could benefit from MR-guidance, as it is prone to anatomical changes. For the purpose of sCT generation, in a first step the required granularity of the image, i.e. the number of required BD overwrite values, was investigated. This was done on the original planning computed tomography (pCT) image to avoid the influence of the choice of the MR sequence, artefacts therein, or the conversion technique. In the second step, the respective BD classes were segmented from an MR image to create MR-based sCT images. The MR image was corrected for common types of artefacts to investigate their impact. The idea behind this approach is described in detail in section 3.3. The quality of the sCT images was analysed in comparison to the pCT based on proton range differences and dosimetric changes. For more details on the image processing and the analysis the reader is referred to section 4.

## 2. Physics background

In this section, the most important principles for magnetic resonance (MR) imaging, computed tomography (CT) imaging and radiotherapy will be explained, which are strongly interlinked nowadays within radiotherapy. CT and MR imaging provide anatomical information about the patient, allowing to identify the tumour volume and critical structures in close proximity to the tumour. In addition, CT provides information necessary for the calculation of energy loss of ionizing radiation in the body. Knowledge on the energy loss is essential for accurately delivering a clinical amount of dose to the tumour, while sparing healthy tissue as much as possible.

Radiotherapy using photons is the most common method, however proton therapy becomes more and more available all over the world. Photons and protons (as charged particles) differ strongly in their physical properties and interaction behaviour when travelling through matter, determining their advantages and disadvantages.

### 2.1. Magnetic Resonance

When protons (spin  $\frac{1}{2}$ ) are placed within a constant magnetic field  $\vec{B}_0$ , they will acquire two spin states  $|\uparrow\rangle$  and  $|\downarrow\rangle$ , named spin up ("parallel") and spin down ("anti-parallel"). The ratio between parallel and anti-parallel spins is given by the Boltzmann distribution, which describes the excess of parallel spins over anti-parallel spins in dependence of the temperature. At body temperature ( $T = 310K$ ) and  $|\vec{B}_0| = 1.5T$  the excess is in the order of  $10^{-6}$ . Considering the huge number of protons in the human body, this excess creates a detectable macroscopic magnetization  $\vec{M}_0$  with magnitude  $M_0$ . Conventionally, the direction of  $\vec{B}_0$  is referred to as the z-direction, giving  $M_0 = M_z$ .

To excite a nuclear magnetic resonance, a time-dependent radio-frequency (RF) pulse is applied irradiating at the Larmor frequency  $\omega_{RF} = \omega_0 = \gamma|\vec{B}_0|$ , with  $\gamma$  being the gyromagnetic ratio. The gyromagnetic ratio is a characteristic property of a nuclei with spin. For protons,  $\gamma \approx 26.75 \text{ radT}^{-1}\text{s}^{-1}$ , leading to a resonance frequency  $\omega_0 \approx 267.51 \text{ MHz}$  or  $\nu_0 \approx 42.58 \text{ MHz}$  at a magnetic field strength of 1 T. The RF field has a polarization such that it acts as a magnetic field  $\vec{B}_1$  perpendicular to  $\vec{B}_0$ . The transmitted pulses are generated by dedicated coils (RF antenna systems) mounted in the bore of the magnet. The magnetization is composed of a component

parallel and a component perpendicular to  $\vec{B}_0$  called  $M_z$  and  $M_{xy}$ , respectively. An RF pulse with a duration such that  $\vec{M}_0$  is flipped by  $90^\circ$  is denoted a  $90^\circ$ - or  $\frac{\pi}{2}$ -pulse. Analogously, any other angle  $\alpha$  can be achieved depending on the length of the pulse. The  $M_{xy}$  magnetization precesses around  $B_0$  with the Larmor frequency and induces a voltage in the RF coils, which is the source of the signal in MR. These receiving coils are generally flexible coils mounted close to the patient for better signal detection.

After RF excitation, relaxation of the transversal ( $M_{xy} \rightarrow 0$ ) and longitudinal ( $M_z \rightarrow M_0$ ) magnetization occurs. It is caused by dipole-dipole interactions between the spins leading to fluctuations in the local magnetic fields and thus changing the Larmor frequency locally (spin-spin relaxation). Also interactions with the environment by thermal coupling to the surrounding material (spin-lattice relaxation) lead to a relaxation. The recovery of the initial state of magnetization is described by the Bloch equations (Bloch, 1946). The spin lattice relaxation induces the recovery of the initial state in z-direction:

$$M_z(t) = M_{z,t=0} \cdot e^{-t/T1} + M_0 \cdot (1 - e^{-t/T1}), \quad (2.1)$$

with  $t = 0$  being the end of the RF pulse, whereas the spin-spin relaxation causes a de-phasing of the individual spin packets, measured as a loss of magnetization in the transversal plane:

$$M_{xy}(t) = M_{xy,t=0} \cdot e^{-i\omega_0 t} \cdot e^{t/T2^*}. \quad (2.2)$$

$T1$  and  $T2/T2^*$  are the longitudinal and the transversal relaxation times, respectively. The magnitude of  $T1$  and  $T2$  depends on the state of the probe: in solids,  $T1$  is very long (hours or days), whereas  $T2$  is very short (milliseconds). In soft human tissue,  $T1$  is in the order of seconds, and  $T2$  is in the order of 100 ms. Thus, different materials can be distinguished based on their relaxation time differences.  $T1$  and  $T2$  are highly dependent on the tissue, which explains why the soft tissue contrast in MR imaging is superior to other imaging methods.

In reality, inhomogeneities of the magnetic field e.g. due to technical limitations or local susceptibility differences, will accelerate the de-phasing of spin packets and thus the decay of  $M_{xy}$ . The overall relaxation time denoted  $T2^*$  is shorter than without these influences ( $T2^* \leq T2$ ). A relaxation with time constant  $T2^*$  directly following the excitation RF pulse is called a free induction decay.

### 2.1.1. The MR Experiment

To distinguish signals from different points in the sample after excitation, additional magnetic fields with well-defined spatial dependence, so-called gradient fields ( $\vec{G}$ ), are superimposed onto  $\vec{B}_0$  such that  $\vec{B} = \vec{B}(\vec{x})$ . This changes the Larmor frequency to be a function of space:

$$\omega_0 = \omega_0(\vec{x}). \quad (2.3)$$

The gradient fields are then used during the MR experiment in the excitation phase for spatially selective excitation, and in the relaxation phase for frequency encoding and phase encoding. Frequency encoding changes the frequency of the magnetization depending on the position, phase encoding will change the phase angle of the magnetization depending on the position. Selective excitation is used for signal differentiation for example in z-direction, then frequency and phase encoding are used for the x-y-plane. In addition, RF pulses of  $180^\circ$  can be used to refocus the decaying signal (caused by the de-phasing of the spins). This way, a second free induction decay (a "spin echo") can be measured at  $TE$  (echo time). A succession of excitation RF pulses, imaging gradients, and refocussing RF pulses is called an MR imaging sequence.

During acquisition of an MR image, the signal intensity at specific points in time along the exponential increase/decay is measured. These points are defined by  $TE$  and the repetition time  $TR$ . Adapting those, focus can be put on differences between tissues in  $T1$ ,  $T2$ , proton density (PD), or combinations of the three. It should however be noted that a PD weighting is present in all images. Depending on which relaxation time is emphasized, the MR sequences are labelled "T1-weighted", "T2-weighted" or "PD-weighted", with the former two being the standard sequences used in the clinics.

The MR images used in this study are mainly T1- and T2-weighted spin echo sequences. Spin echo sequences have been developed in the early years of MR imaging and are still widely used. The basic pulse sequence is  $[90^\circ - T_E/2 - 180^\circ - T_E/2 - \text{AQ} - T_D]$  with AQ being the signal acquisition and  $T_D = T_R - T_E$  the delay time until the next pulse sequence starts. The signal intensity  $S_{SE}$  of a voxel is given by

$$S_{SE} = PD \cdot [1 - e^{-TR/T1}] \cdot e^{-TE/T2}. \quad (2.4)$$

T1-weighting can be achieved with  $T_R = T1$  and  $T_E \ll T2$ , T2-weighting with  $T_R \gg T1$  and  $T_E = T2$ .

Gradients for slice selection and phase and frequency encoding are applied sequentially, the latter two are varied N, respectively M times, to fill a plane with signal intensities in the frequency domain. Generally, N is equal to M, to obtain quadratic pixels in the image and a quadratic image. To obtain the MR image, a two-dimensional (2D) Fourier transform is applied. The resulting image also has a resolution of NxM pixels. Thus, in this simple form, the basic sequence has to be repeated NxMxP times, to achieve a three-dimensional (3D) image with P slices. Fast spin echo sequences use multiple echoes instead of just one echo with varying phase encoding to decrease acquisition time. An example of the gradient sequence for a fast spin echo sequence is shown in figure 2.1.

Examples of T1- and T2-weighted MR images acquired with fast spin echo sequences are shown in figure 2.2.

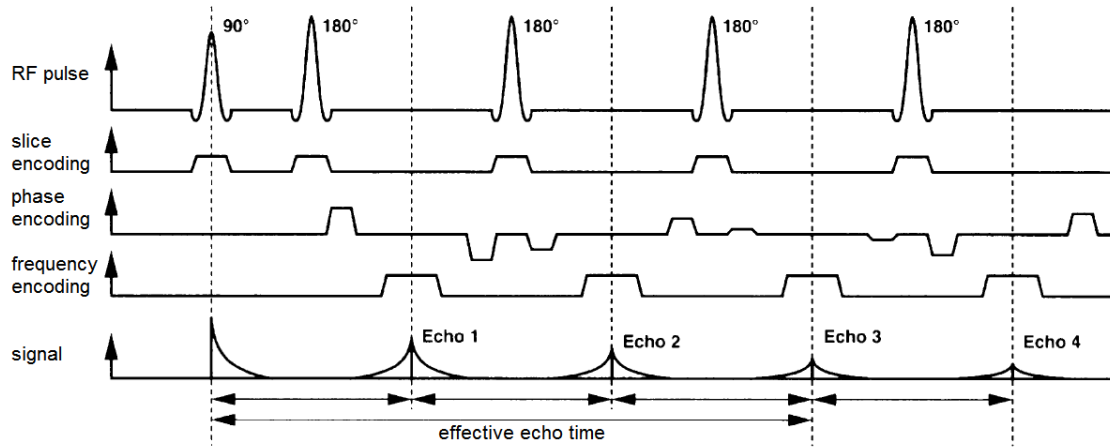
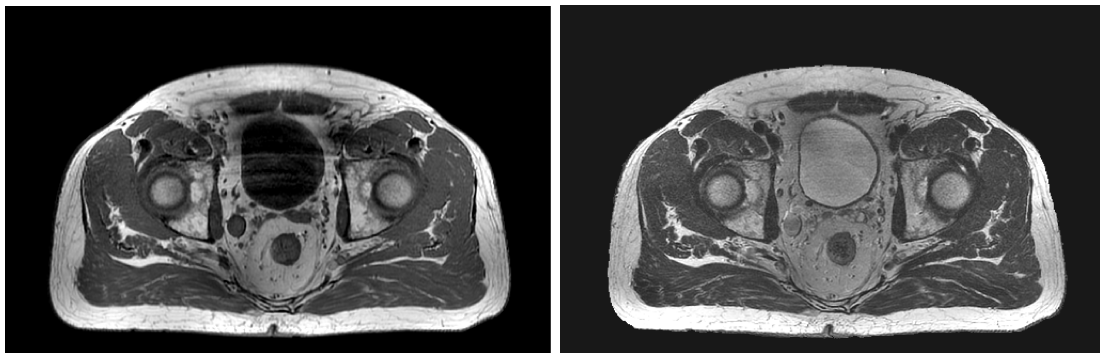


Figure 2.1.: Gradient sequence for a fast spin echo sequence with four echos, i.e. four  $180^\circ$ -pulses. Figure adapted from Weishaupt et al. (2014)



(a) T1-weighted MR image ( $T_E = 7.2\text{ms}$  and  $T_R = 600\text{ms}$ ). (b) T2-weighted MR image ( $T_E = 97\text{ms}$  and  $T_R = 6000\text{ms}$ ).

Figure 2.2.: Two images acquired with fast spin echo sequences showing two different contrasts depending on the echo time  $T_E$  and repetition time  $T_R$ .



### 2.1.2. Artefacts in MR

MR imaging is a complex process, both physically and technically, giving rise to many artefacts. Here, artefacts relevant for dose calculation based on MR images are briefly explained.

1. The air/bone ambiguity

In T2-weighted MR sequences, both bone and gas, for example air, have very low signal intensity. In air, the reason is the low density of protons that contribute to the macroscopic magnetization and thus the signal. In bone, enough protons are available, but as a solid material, the decay of the transversal magnetization is too fast to detect the signal of an echo with standard echo times. Therefore, sequences with ultra short echo time (UTE) or zero echo time (ZTE) (which simply measure the free induction decay) have been developed. Bone will then have a small signal, whereas air is still void.

2. Image distortion

The localisation of signal in MR is done via gradient fields overlaid on a homogeneous, static field  $\vec{B}_0$ . If either  $\vec{B}_0$  is not perfectly homogeneous, or the gradient fields are not exactly linear, the detected signal will appear to originate from a slightly different location than expected. In the image, this can be seen as distortions. Also the patient itself can affect  $\vec{B}_0$  via susceptibility effects, and thus introduce inhomogeneities in the magnetic fields, leading to distortions.

3. Bias field

A bias field is a slowly varying, additive field, which may appear to the observer as an uneven illumination of the image. In figure 2.2, it is clear to see at the subcutaneous fat in both images, which from a physical point of view should have the same grey value independent of location in the image. However, some areas are brighter than others. This effect is caused by inhomogeneities in the  $\vec{B}_1$  field and inhomogeneity of the sensitivity of the RF coil that receives the MR signal. Generally, regions close to the coil appear brighter (higher sensitivity of the coil).

## 2.2. Computed Tomography

CT imaging refers to a 3D image of a patient based on the transmission of x-rays through the body. During transmission, the photons interact with the atoms in the body, changing the intensity of the in-going photon beam according to Lambert-Beer's law of absorption:

$$T(x, E_\gamma, Z) = \frac{I(x)}{I(0)} = e^{-\mu(E_\gamma, Z) \cdot x} \quad (2.5)$$

with  $T$  the transmission,  $I$  the beam intensity,  $x$  the traversed path,  $E_\gamma$  the energy of the photons,  $Z$  the effective atomic charge number of the traversed material, and  $\mu$  the linear absorption coefficient of the material.

Three physical effects contribute to the absorption coefficient depending on the beam's energy and the charge number: the photoelectric effect (PE), Compton scattering (CS) and pair production (PP). The total energy of the incoming photon in the PE is used to emit an electron from a target atom, which will deposit its energy locally. The absorption coefficient of the PE,  $\mu_{PE}$ , is approximately proportional to  $Z^3/E_\gamma^3$ . Figure 2.3 illustrates that the PE dominates the absorption coefficient (here normalized to mass density) for photon energies below 30keV and is generally only relevant below 100keV. Above that, CS is the most important effect. CS is the incoherent scattering of the photon at an electron, which will then be emitted from the atom. From figure 2.3 it can also be seen that CS, especially in the energy regime relevant for CT imaging (30-150keV), is almost independent from the photon energy. It is also hardly dependent on  $Z$ , but shows proportionality to the electron density  $\rho_e$  of the absorbing material. PP of an electron-positron pair in the strong electromagnetic field of an atomic nuclei can only happen at photon energies above a threshold of 1.022MeV. The positron will annihilate with an electron of the material and thereby emit two photons with  $E = 0.511\text{MeV}$ . Only above 20MeV, the impact of PP is dominant, as can be seen in figure 2.3.  $\mu_{PP}$  is proportional to  $Z$  and increases with energy. PP is not relevant for imaging, because the photon energies are well below the threshold.

Equation 2.5 assumes a beam passing through a homogeneous material. The human body however is very heterogeneous in its composition ( $Z$  and  $\rho_e$ ), and thus are the absorption coefficients. This changes equation 2.5 as follows:

$$T(x, E_\gamma, Z) = e^{-\int_0^x \mu(E_\gamma, Z, x) dx}. \quad (2.6)$$

In CT imaging, absorption profiles through the body are acquired from different directions (covering at least  $180^\circ$ ), as shown in figure 2.4 (left) for the first CT scanner designs. After scanning many parallel lines to create one absorption profile, the x-ray tube and the scanner are rotated and the measurement starts again. At this point, the absorption coefficient of each voxel in the field-of-view (FOV) is still unknown, but contributes to each absorption line that passes the voxel. With that, an equation system is created, whose solution gives the CT image. Nowadays, the most common solutions are analytical and iterative reconstruction algorithms.

This implementation for measuring absorption lines is very slow, which is why nowadays all clinical scanners consist of an x-ray tube creating a fan-shaped beam (cf. figure 2.4 (right)), mounted on a gantry ring. On the gantry, a curved, pixelated solid-state detector is mounted opposite of the x-ray tube. This set-up is called fan-beam geometry. Thus, many absorption lines are measured simultaneously instead of successively. To scan a slice, the gantry rotates around the patient. Modern

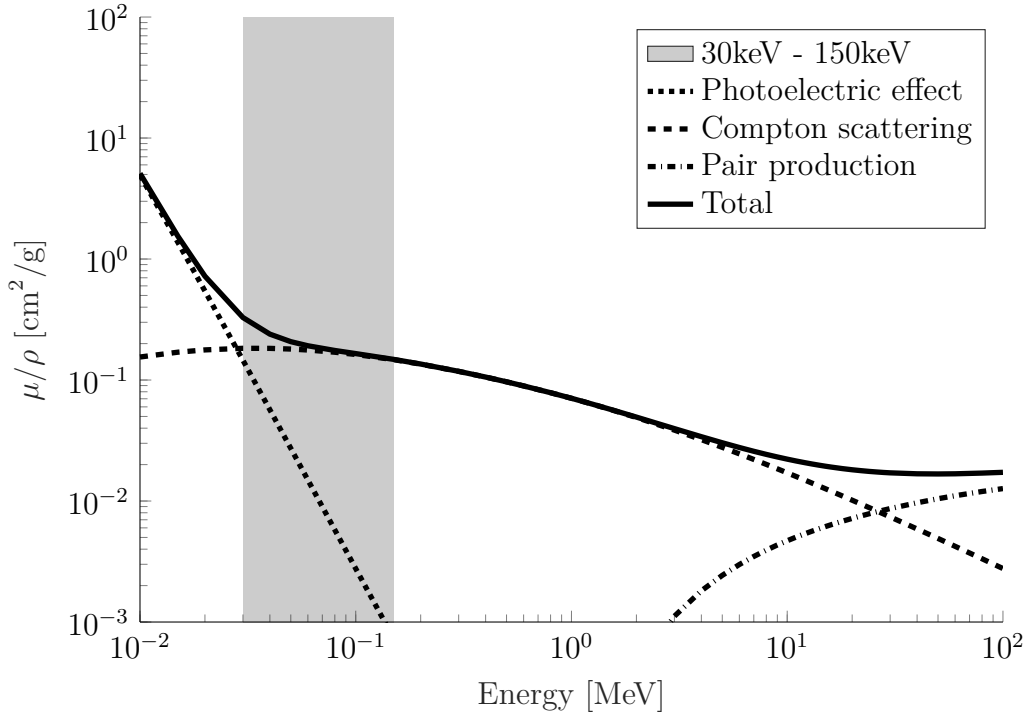


Figure 2.3.: Absorption coefficients of water normalized by the mass density over the photon energy for the photoelectric effect, Compton scattering, pair production and the total sum of them. The grey area indicates the energy range relevant for CT imaging (30keV - 150 keV). Data taken from the NIST XCOM Photon Cross Section Database (Berger et al.)

scanners also have several rows of detectors to measure more than just one slice at a time. For faster volume acquisition, the gantry rotation is overlaid to a movement in z-direction (defined as the cranial-caudal direction), i.e. a spiral movement, instead of acquiring one slice at a time.

In a CT image, the grey value of a voxel represents the spatial distribution of absorption coefficients of the material within the voxel. To make images from different scanners which may have different energy spectra comparable, a normalization of the absorption coefficients to water is introduced, called the CT scale given in Hounsfield unit (HU):

$$H(\mu) = \frac{\mu - \mu_w}{\mu_w} \cdot 1000 \quad [\text{HU}]. \quad (2.7)$$

$\mu$  is the absorption coefficient of the voxel, and  $\mu_w$  the absorption coefficient of water. The scale ranges from -1000 HU for vacuum to 0 for water and continues theoretically to infinity. In practice, it ranges from -1024 HU for vacuum to 3071 HU for high absorption material, as it can be stored in binary system with only 12 digits. Soft tissues range between -120 HU for fat and 300 HU for muscles and organs. Bony tissue ranges from 300 to 1900 HU. Foreign materials, e.g. metals in prosthetics, are at the upper end of the scale and clipped to 3071 HU.

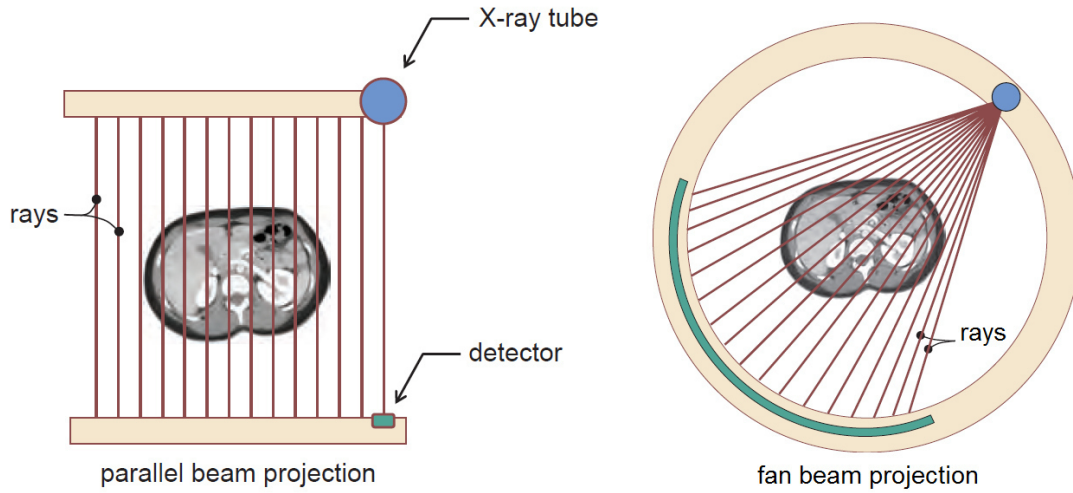


Figure 2.4.: Schematics of CT scanning methods. On the left, the first generation of scanners with a pencil-beam scanning parallel lines before rotating to a different angle is shown. On the right, an x-ray tube which generates a so-called fan-beam is shown, which is detected by a curved, pixelated detector. Image source: Bushberg et al. (2012).

One disadvantages of CT imaging are the poor image contrast within soft tissues, which are for example necessary in oncology to locate tumours and metastases. Also the dose burden to the patient from the x-rays might hinder more frequent use of CT imaging. On the other hand, the image contrast between soft tissue and bone is very high, and 3D images with high spatial accuracy can be acquired much faster than with other imaging modalities. The spacial accuracy is ensured by the application of projections, instead of using a mediating property for voxel locations as the magnetic field in the case of MR imaging. The most important property of CT imaging in the context of oncology however is its direct dependency on electron density and atomic number of the tissue, which are the relevant parameters for dose calculation in radiotherapy.

## 2.3. Radiotherapy

Radiotherapy is one of the main pillars of cancer therapy, together with surgery and chemotherapy. Its principle is to deliver a high amount of energy to the tumour cells, which will ultimately kill the cells, while trying to spare healthy tissue. The commonly used quantity to describe this delivered energy is dose, which is defined as the mean absorbed energy per mass:

$$D = \frac{d\overline{E}}{dm} \quad [\text{J kg}^{-1} = \text{Gy}] \quad (2.8)$$

When physicians decide to apply radiotherapy for cancer therapy of a patient, several steps have to be taken before irradiating the tumour. First, a computer model of the patient is created. This model is based on the CT image of the patient in treatment position, which has high spatial accuracy and contains information on the energy absorption properties of the tissue. MR images or functional images like positron emission tomography (PET) can add information on soft tissue and tumour location, which might not be available in the CT. These images have to be registered to the CT for spatial correlation. In a second step, the gross tumour volume (GTV) and organs at risk (OARs), i.e. healthy organs that could suffer irreparable damage from radiation, are contoured in the images. The GTV will be extended to include tumour spread at the cell-level to form a clinical target volume (CTV). To include planning and treatment uncertainties, the CTV is further extended to form a planning target volume (PTV). The additional volume added by the PTV acts as a safety margin.

Afterwards, the dose that should be delivered to the PTV is prescribed and dose constraints for OARs are assigned. Dose constraints are well-defined thresholds, that should not be exceeded. For example: the maximum dose considering all voxels in the urinary bladder should not exceed 65 Gy (Marks et al., 2010). Dose constraints are generally based on dose-volume histograms (explained in detail in section 4.7). Prescribed dose and dose constraints are important input parameters for the calculation of the treatment plan in the next step. Nowadays, only limited parameters like treatment particle, number of beam angles, tumour volume, etc., are set beforehand. Other parameters, like the weighting of individual pencil-beams within one beam, are optimized in an iterative process called inverse planning. Only after all requirements are met, i.e. the tumour receives the prescribed dose and all OARs are spared according to the constraints, the treatment is delivered to the patient.

Another method to spare healthy tissue is to deliver the treatment in a fractionated fashion. Instead of delivering all dose in one session, it is split into smaller doses which are delivered successively on different days. This leads to a higher tumour control while lowering the risk of permanent damage in healthy tissue. Typically, a treatment is delivered in 20 - 30 fractions with 1.8 - 2 Gy each, but other schemes are possible depending on the indication.

### 2.3.1. Intensity modulated radiotherapy

The most common type of radiotherapy is using high energetic photons (approximately between 1 and 10 MeV), sometimes referred to as "conventional radiotherapy". Simple x-ray tubes are not sufficient any more to produce such high energy photons. Instead, electrons are accelerated using a linear accelerator (linac) to reach respective energies that create ultra-hard x-rays when hitting a metal target. To describe the beam's energy, the accelerating voltage of the electrons is given in MV

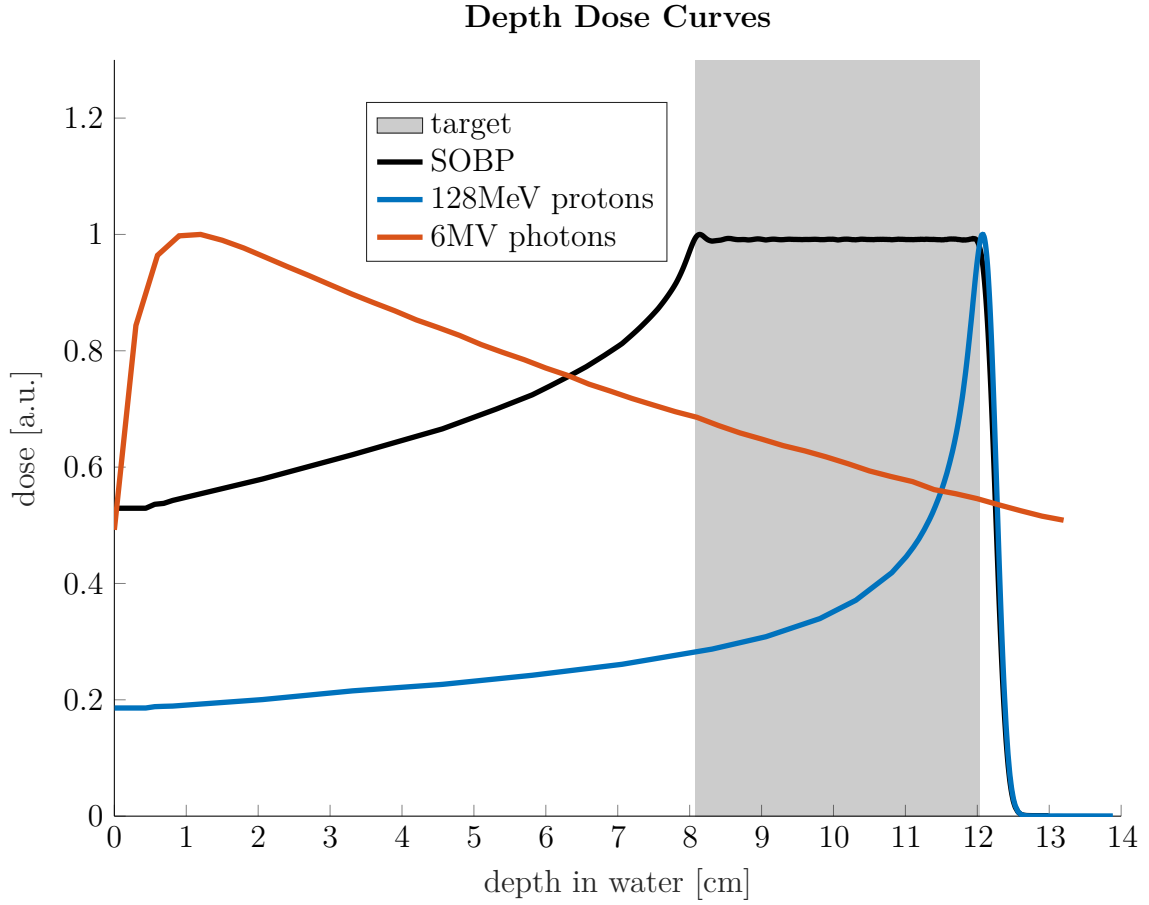


Figure 2.5.: Depth dose curves produced by a 6MV photon beam (orange), by a 128MeV proton beam (blue), and by a spread-out Bragg peak (SOBP) (black) covering a target (grey area).

rather than the energy of a monoenergetic beam, because the output of the linac is a spectrum of energies. The resulting radiation is collimated to form a beam. The physics of the photons in the beam hitting matter is similar to the physics in CT imaging, however the higher energies have the consequence that the absorption is dominated by CS and, above 1.022 MeV, also PP, whereas the PE can be neglected (cf. figure 2.3). The energy deposition in the patient mostly occurs due to the secondary electrons from the CS in Coulomb interactions with electrons of the body. Thus, a build up effect at the beam's entrance into the patient can be observed, which reduces the dose to the skin. After the build-up, secondary electron equilibrium can be assumed, in which the dose deposition decreases exponentially following Lambert-Beer's law while traversing matter, as seen for the orange curve for a 6 MV photon beam in figure 2.5.

For dose calculation, a fast analytical approach, which is widely used, is the pencil-beam algorithm (Bortfeld et al., 1993). It calculates the dose at a point  $(x_P, y_P, z_{rad})$

deposited by a pencil-beam with the central axis at  $(x, y)$  as

$$D(x_P, y_P, z_{rad}) = \iint_{-\infty}^{\infty} \Phi(x, y) F(x, y) K(x - x_P, y - y_P, z_{rad}) dx dy. \quad (2.9)$$

$\Phi(x, y)$  is the primary fluence of the beam,  $F(x, y)$  the transmission factor transcribing the intensity-modulation of the beam (i.e. a collimation of the beam to the projected shape of the tumour) and  $K$  is 3D kernel, which describes the shape of the beam laterally and in depth.  $z_{rad}$  is the radiological depth (rD) that determines the absorbed dose in a heterogeneous material (also compare equation 2.6):

$$z_{rad} = \int_0^P \mu(l, E_\gamma) dl. \quad (2.10)$$

The absorption information  $\mu$  is derived from the CT image using a Hounsfield look-up table (HLUT) (for more details see below), which has discrete voxel information. Thus, the rD is not an integral but a sum over the products of the length of the path within voxel  $i$ ,  $l_i$ , and the absorption coefficient in this voxel,  $\mu_i$ :

$$z_{rad} = \sum_{i=0}^N l_i \mu_i. \quad (2.11)$$

To characterize the kernel, the dose calculation algorithm depends on detailed knowledge of the pencil-beams, which are measured or simulated in water in advance. For this reason, it is convenient to make all calculations "relative to water". The rD relative to water is

$$\hat{z}_{rad} = \frac{z_{rad}}{z_{rad,w}} = \sum_{i=0}^N l_i \frac{\mu_i}{\mu_w} = \sum_{i=0}^N l_i \frac{\rho_{e,i}}{\rho_{e,w}} = \sum_{i=0}^N l_i \hat{\rho}_{e,i}. \quad (2.12)$$

$\hat{\rho}_{e,i}$  is the relative electron density (rED), which is the dominant tissue property to affect the absorption of MeV-photons. The calculation of  $\hat{z}_{rad}$  for all target points in the tumour volume is called ray casting.

Thus, it is assumed that spatial rED information of the patient is available. But the energy dependence of the absorption behaviour of tissue (cf. figure 2.3) implies that the CT numbers can not directly be used for dose calculation. Instead, the CT image is translated to a rED image using the so-called HLUT. The HLUT is a piece-wise linear function and determined based on measurements of known materials and/or tissue surrogates.

With a pencil-beam algorithm, a calculation of dose is possible given its energy, fluence and transmission factor. This is called forward calculation. To reach a prescribed dose in the tumour while sparing healthy tissue, several beams are irradiated from different angles, overlapping in the target (as illustrated in figure 2.6 (left)). For complex geometries, for example when an OAR lies in close proximity to the tumour, it is often difficult to spare the OARs appropriately. In this case, the intensity across the beam can be adapted, which is advantageous for example for concave

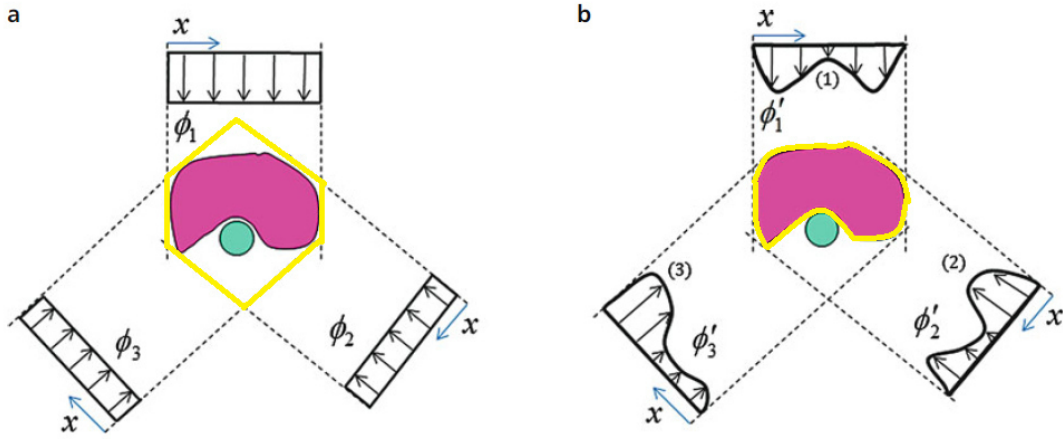


Figure 2.6.: Comparison between non-modulated (left) and intensity-modulated radiotherapy (IMRT, right). In IMRT the intensity of the primary beam  $\Phi$  is adapted across the beam to deliver lower doses to OARs, as the green structure in the example shown here. The high dose area delivered by the two methods is indicated by the yellow line each. Adapted from: Schlegel et al. (2018).

tumour shapes around an OAR. This technique is referred to as intensity-modulated radiotherapy (IMRT). The difference between the non-modulated radiotherapy and IMRT is demonstrated in figure 2.6: on the left, the OAR in green receives the same dose as the tumour in red indicated by the yellow line for the high dose region. On the right, the intensity of the beam is reduced where the beam passes through the OAR and increased where the beam only passes through tumour.

The variation of the intensity within one beam increases the degrees of freedom for the plan parameters. It is not feasible to manually set all parameters and do a forward calculation. Thus, only the number of beams and their angles are specified by experienced physicians. The shape of the beams is defined by the projected contour of the tumour. This leaves only the intensity modulation across the beams to be optimized by computational methods before the dose can be calculated. Therefore, this approach is called "inverse treatment planning". The optimization is based on a linear relationship between the dose in a voxel  $i$  and the intensity of a partial fluence element of the beam, or pencil-beam, called "bixel"  $j$ :

$$d_i = \sum_j D_{ij} w_j, \quad (2.13)$$

with  $D_{ij}$  being the so-called dose influence matrix. The intensity vector  $w_j$ , which describes the intensity modulation of a beam, can be optimized iteratively based on objective functions  $F$ , that penalize for example a dose difference from the prescribed dose in the tumour, or exceeding of a dose constraint in an OAR. The solution of the inverse problem is the minimization of the objective function:

$$w^* = \operatorname{argmin}_{w \geq 0} F(w).$$



### 2.3.2. Intensity modulated proton therapy

Proton therapy, another type of radiotherapy, is getting more and more available worldwide. Protons are charged particles, that interact via electromagnetic and hadronic forces with matter. The mean energy loss  $d\overline{E}$  during these interactions on the path  $dx$  is described by the stopping power:

$$S = \frac{d\overline{E}}{dx}. \quad (2.14)$$

Hadronic forces, i.e. interactions with atomic nuclei of the traversed matter, lead to fragmentations of the nuclei. Also elastic interactions with atomic nuclei occur, however the energy loss is negligible in these interactions, i.e. the stopping power  $S_{nuc}$  does not contribute significantly to the overall stopping of the protons. The most important interaction is the Coulomb interaction with electrons also known as the collision stopping power  $S_{col}$ , which slows down the protons continuously. This interaction is dominating the dose:

$$D \sim \frac{S_{col}}{\rho} \cdot \Phi \quad (2.15)$$

with  $\Phi$  the primary fluence and  $\rho$  the mass density. The fraction  $S_{col}/\rho$  is described by the Bethe formula (Bethe, 1930):

$$\frac{S_{col}}{\rho} = k \cdot \frac{Z_A}{A_A} \cdot \frac{1}{\beta^2} \cdot \left[ \frac{1}{2} \ln \left( \frac{2m_e c^2 \gamma^2 \beta^2}{I} \right) - \beta^2 \right]. \quad (2.16)$$

$k$  is a constant factor ( $\approx 0.307 \text{ MeV} \cdot \text{cm}^2/\text{g}$ ),  $Z_A$  the charge number,  $A_A$  the atomic number of the traversed matter,  $\beta = v/c$  the velocity,  $m_e$  the mass of an electron,  $c$  the speed of light,  $\gamma$  is defined as  $1/\sqrt{1-\beta^2}$ , and  $I$  is the mean excitation energy of the traversed material. Correction factors for high and low energies are neglected here. In analogy to the rED, the relative stopping power to water, also called stopping power ratio (SPR), is introduced as

$$SPR = \frac{S_{matter}}{S_{water}}. \quad (2.17)$$

The range of the particle entering matter with energy  $E$  under the continuous slowing down approximation is defined as

$$R(E) = \frac{1}{\rho} \int_0^E \frac{1}{S(E') dE'}. \quad (2.18)$$

A proton with energy 220 MeV for example has a range of approximately 30 cm in water.

The characteristic stopping behaviour of charged particles ( $-dE/dx \propto 1/\beta^2$ ) leads to an advantageous dose deposition in the patient: low dose in the entrance channel

and high dose in the so-called Bragg peak at the end of the particles range. Figure 2.5 shows a single Bragg peak produced by a 128 MeV proton beam (blue line). To cover an extended tumour volume in depth (grey area in the figure), several Bragg peaks from protons with different energies are overlaid to form a spread-out Bragg peak (SOBP), as shown as a black line in figure 2.5. To cover the tumour in lateral directions, several pencil-beams are placed next to each other. The sum of all pencil-beams from the same entrance direction is called a beam. The intensity, i.e. the number of protons, can be varied between pencil-beams within a beam, and the modulation of the intensity is likewise advantageous as in photon therapy to spare OARs. This technique was therefore introduced as intensity-modulated proton therapy (IMPT).

The fastest way to calculate the dose of a proton beam, similar to photon dose calculation, is a pencil-beam algorithm (Hong et al., 1996). The description of a pencil-beam is split into a lateral dose profile and a depth dose curve:

$$d(\vec{r}) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{(x-\mu_x)^2}{2\sigma^2}} \cdot e^{-\frac{(y-\mu_y)^2}{2\sigma^2}} \cdot Z(z_{rad}), \quad (2.19)$$

with  $\mu_x$  and  $\mu_y$  are the lateral beam position. From equation 2.19 it can be seen that the lateral dose profile can be described with a Gaussian function with beam width  $\sigma$ .  $\sigma$  also has a dependency on  $z_{rad}$  as a consequence of multiple Coulomb scattering (MCS) and inelastic interactions, thus both  $\sigma$  and the depth dose curve  $Z(z_{rad})$  are based on measurements in water. The rD is defined in analogy to photon therapy:

$$z_{rad} = \sum_{i=0}^N l_i SPR. \quad (2.20)$$

The process of the treatment planning is also comparable to IMRT. First, the CT image is converted using a HLUT to SPR. The HLUT is generated via measurements of CT numbers and SPR using known materials and/or tissue surrogates. As for the photon HLUT, the proton HLUT is a piece-wise linear function.

Compared to photon therapy, less beams are sufficient to cover the target adequately while sparing the healthy tissue sufficiently. Typically, 2-3 beams are used in IMPT, compared to 6-9 beams in IMRT. The reason in the inverse dose profile of the proton beam, as it was visualized for a SOBP in figure 2.5. Less beams ultimately also lead to a lower total dose in healthy tissue, as the beams go through less volume.

When delivering the same physical dose to cells with photons and protons, the fraction of surviving cells will be lower for proton therapy. The reason is the higher efficiency of protons to cause lethal complex DNA breaks in the cells. For treatment planning, this effect is considered using the concept of relative biological effectiveness (RBE). RBE is defined as the ratio of dose delivered via photons and dose delivered via protons, leading to the same iso-effect. For protons, a constant RBE of 1.1 is

used for dose calculation. To distinguish whether a given dose is physical dose or biologically effective dose, it is reported as Gy or Gy(RBE), respectively.

Ray casting is performed to know the rD of each target point inside the tumour, and the dose influence matrix is calculated. As the dose optimization has an additional degree of freedom compared to photon therapy, namely the depth modulation using proton beams with different energies, the  $D_{ij}$  matrix is much larger than in the photon case, where the energy is fixed. Dose prescriptions and constraints are included in the objective functions  $F$ , which are iteratively minimized to find the optimal intensity vector  $w_j$ .

### 2.3.3. Uncertainties in radiotherapy

The complexity of radiotherapy causes many sources of uncertainties. Describing and explaining all uncertainties and their origin is beyond the scope of this thesis, thus only uncertainties relevant to the analysis performed during this work are explained.

CT imaging offers high spatial accuracy with absorption information. To overcome its low tissue contrast, MR and if necessary also PET images are acquired and used for delineation. Nevertheless, the delineation is done manually and introduces an intra- and inter-observer variability. Fiorino et al. (1998b) performed a comparison study with five trained radiotherapists, that contoured the prostate in CT images of six patients. The intra- and inter-observer variability in prostate volume was up to 9 % and 18 %, respectively. It can be expected that the variability decreases when using MR images. These images, however, introduce further errors. Firstly, the MR is prone to distortions. Even after distortion correction, the prostate volume can suffer from distortions up to 0.5 mm (Nyholm et al., 2009). Secondly, the contours have to be propagated to the CT, on which further treatment planning is conducted. This propagation is done via image registration, which introduces an error of approximately 2 mm (Roberson et al., 2005).

Range uncertainty affects both treatment modalities, however the impact in proton therapy is more severe due to the steep dose gradients compared to photon therapy. One source of range uncertainty is the applied HLUT, which cannot cover non-tissue-like materials such as implants (Schaffner and Pedroni, 1998) or inter-patient variations (Woodard and White, 1986). So far, also no real patient tissue has been used to create HLUTs, which introduces the uncertainty of the surrogates themselves. The reported uncertainty introduced solely by the HLUT is about 1 % of proton range (Schaffner and Pedroni, 1998).

After treatment planning, the largest uncertainties are related to patient positioning and inter- and intra-fractional anatomical variations (Lomax, 2008). Correct

positioning of the patient is essential for the accuracy of the irradiation, and measures, such as thermoplastic masks for head irradiations or special mattresses for irradiations in the abdominal or pelvic region, are taken to make patient positioning reliable. Still, uncertainties of about 3 mm have been reported by Fiorino et al. (1998a). Pre-treatment imaging, for example using x-ray imaging or cone-beam CT (CBCT), can be used to detect and correct misalignment of the patient, and is based on reference points such as fiducial markers or bony anatomy. Wu et al. (2001) reported random and systematic isocentre position shifts of 2.2 mm and 1.4 mm, respectively. This, however, does not reflect and correct intra- and inter-fractional motion relative to the reference points. Intra- and inter-fractional motion can include heart beat, breathing, organ motion, weight loss or tumour shrinkage. For prostate cancer for example, organ motion can be up to 10mm (Roeske et al., 1995), not including other anatomical variations.

## 3. MR-based dose calculation

The following section explains the rationale behind dose calculation on magnetic resonance (MR) images. Several methods that are currently available in the literature are explained briefly detailing their advantages and disadvantages. The goal of this thesis and the idea behind its implementation is described at the end of this section.

### 3.1. Motivation

From a physical point of view, dose calculation on MR images for radiotherapy is neither obvious nor reasonable: the properties of the tissue that are probed in the MR are not related to the necessary properties needed for dose calculation (either relative electron density (rED) or stopping power ratio (SPR)). Also, MR is prone to artefacts such as distortions induced by magnetic field inhomogeneities, patient-induced inhomogeneities, or gradient non-linearities (Tanner et al., 2000), which can only partially be accounted and corrected for. In comparison, the geometrical fidelity is very high in computed tomography (CT) imaging. An advantage of the MR is that it provides images with superior soft tissue contrast (Austin-Seymour et al., 1995) and is already used in the context of radiotherapy for volume delineations (ICRU, 2010). These delineations, however, still have to be transferred to the CT via image registration, which introduces an error to the resulting treatment plan (Nyholm et al., 2009). MR-only treatment planning could avoid this registration error and spare the effort of imaging both CT and MR. Another advantage of MR imaging is the absence of ionizing radiation, as compared to CT imaging with x-ray radiation. The patient is not exposed to an additional dose burden due to imaging. Besides, frequent imaging, which would potentiate the dose burden, is especially important for tumours that shift inter- or intra-fractional. Examples for such shifts can be organ motion or positional shifts of the patient on the treatment couch. Imaging is also important in cases where the tumour is in close proximity to critical organs at risk (OARs), where exact knowledge on position of the tumour relative to the OAR is crucial for treatment outcome. For this reason, investigations on how to make use of the MR image for dose calculation have already started around the year 2000 (Khoo et al., 1997; Beavis et al., 1998; Lee et al., 2003; Chen et al., 2004) and have increased steadily over the past years. The only investigated approach so far is to generate so-called synthetic computed tomographies (sCTs) images from

the MR image. Independent from the method used to generate sCT images, all methods detect certain properties of the tissue in the MR image (e.g. grey value, location of voxels, texture, etc.) and assign CT numbers to that tissue. An overview over the most common approaches is given in the following section. A boost in MR technology, such as new MR sequences, and an increase in computational power have accelerated the use of MR imaging steadily over the past years.

## 3.2. Available Methods

The algorithms reported in the literature to generate sCT images can be divided by their approaches into three categories: bulk density (BD) overwrites, voxel-based and atlas-based algorithms.

BD algorithms are the simplest and most robust algorithms. They have been the first algorithms investigated for sCT image generation, beginning with complete water-equivalent (WE) overwrites of the patient anatomy (Schad et al., 1994; Chen et al., 2004). Those approaches are very fast, since only the patient body contour has to be extracted from the MR, however it is only sensitive to body contour changes and not internal anatomical changes. They can also not be used for position verification, which relies on the detection of bony anatomy and comparing it to a reference. Therefore, several studies added a BD value for (mainly manually contoured) bones (Lee et al., 2003; Eilertsen et al., 2008; Jonsson et al., 2010). These studies showed better dosimetric agreement with the planning computed tomography (pCT) and enabled position verification. But manual contouring is too time-consuming for clinical routine, especially in the head and neck region. Besides bones, also air cavities have significant impact on dose distributions. They might be added to the sCT (Korsholm et al., 2014), but are hard to distinguish with standard clinical MR sequences from cortical bone without manual contouring done. Most commercial solutions for MR-based dose calculation for photon radiotherapy (as for example integrated into combined radiotherapy and MR machines, so-called "MR-linear accelerator (linac)") are applying BD methods, as described for example in Tyagi et al. (2017).

From this approach, sCT generation developed into two directions. One is atlas-based, which uses the information from a library of CT and MR image pairs (the "atlas") to generate sCT images from MR, i.e. by deformation of atlas-CTs to the anatomy of the incoming MR (Uh et al. (2014); Demol et al. (2016); Kraus et al. (2017); a commercial solution suggested by Siversson et al. (2015)). Another, related method is to find small areas with similar structure(so-called "patches") the incoming MR and in the atlas-MR, and then replace them patch-wise with corresponding patches from the atlas-CT (Andreasen et al., 2015). This generates very good sCT images, however these methods are limited by the time-consumption and quality of the deformation and can only be as good as the atlas data. For

patients with anatomic anomalies, for example after surgery, the algorithms produce artefacts hindering the usability of the sCT images.

As an alternative, voxel-based methods have been developed, which are not limited to standard anatomies, but translate the MR to a sCT mostly using machine learning methods, but also simpler translation-functions. One method that has been evaluated in many studies, the dual model by Kapanen and Tenhunen (2012), applies a polynomial model for bones and a regression model for soft tissues, extracted from a single MR sequence. In the beginning, bones had to be segmented by hand, because grey value differentiation was not feasible between air and cortical bone, until an automated segmentation (Koivula et al., 2016) was introduced.

The ambiguity between air and bone was one of the driving factors for the advance of special sequences for MR imaging. Cortical bone has very short T2 relaxation times, and can therefore not be distinguished from the signal-less air. Sequences with ultra short echo time (UTE) or zero echo time (ZTE) have therefore been extensively investigated for sCT generation to measure signal from cortical bone, often in combination with other MR sequences for better soft tissue separation (Johansson et al., 2011; Hsu et al., 2013; Rank et al., 2013; Zheng et al., 2015; Wiesinger et al., 2018). It should be mentioned that these additional sequences add more scanning time for the patient in the MR, which increases the costs and the chances for anatomical changes between the scans.

The use of multiple MR sequence images initiated also more advanced sCT generation methods: machine learning methods of different kinds have been applied in several studies. Rank et al. (2013) and Hsu et al. (2013) used classification approaches covering the whole Hounsfield scale using two MR images as input, Johansson et al. (2011) applied a regression model.

Recently, a trend towards neural networks can be observed among the published studies (Han, 2017; Maspero et al., 2018; Chen et al., 2018; Spadea et al., 2019). Trained neural networks can generate sCT from incoming MR images very fast without the need of multiple and/or special sequences, however require ideally large training datasets with high anatomical agreement between CT and MR images.

Wiesinger et al. (2018) were the first to use (a single) ZTE image, which allowed the use of a simple linear function and thresholding to generate sCT images from the MR. The UTE/ZTE images improve the separation of air and bone, however those sequences are no clinical standard, thus need extra time for acquisition in addition to the clinical standard sequences.

### 3.3. Goal of this thesis

Ideally, an algorithm to generate sCT images is fast, robust, compatible with different scanners/MR sequences at different clinics, and accurate enough to detect position or anatomical changes, which propagate to dosimetric changes relative to

the pCT. Proton therapy might especially benefit from MR imaging, because even small changes in position or anatomy can lead to severe changes in dose deposition due to the steep dose gradients. So far, machine learning algorithms (Rank et al., 2013; Edmund et al., 2014; Zheng et al., 2015; Pileggi et al., 2018; Spadea et al., 2019) have been applied for proton therapy, or existing algorithms were adapted (Koivula et al., 2016; Maspero et al., 2017). Compared to photon therapy, there has been no systematic increase in complexity of the algorithms as it could be expected from the increased susceptibility to changes.

At the same time, it is not clear whether more complex algorithms are necessary for the application of sCT images. It cannot easily be estimated if an increase of algorithm complexity is equivalent to an increase in quality of the sCT. However, the granularity of the sCT as a substitute for complexity can be investigated. Granularity depicts the number of BDs in the sCT image. The lowest granularity has a sCT image with one BD, as for example a WE image. The highest granularity covers the full range of CT numbers, as all voxel-based algorithms do.

In this thesis, the necessary granularity for sCT images in proton therapy is investigated. The investigation is performed in two steps:

1. The first step reduces the granularity of pCT images without the use of MR images. This procedure allows to identify tissue classes, that have a significant impact on the dose distribution, and that should be identified in the MR image. The idea is to replace different segments of the pCT with BD values, starting from one BD value for the whole body and then gradually increase the number of BDs corresponding to different classes of tissue. The less BDs an image has, the more robust is the image and the algorithm to create it. The direct, voxel-wise comparison of such an bulk density computed tomography (BD-CT) to the pCT, for example using the mean absolute error (MAE) as a metric (cf. 4.5.1), will naturally be poor. For this reason, the quality of the BD-CT is evaluated based on re-calculations of the dose distribution (DD), which has been optimized using the pCT as the gold standard.
2. Based on the results of the granularity investigation, respective tissue classes are segmented from the MR image and replaced with BD values. Both the original MR image, and the deformably registered MR image were used as input for sCT generation. Neglecting distortions of the MR image, this allows to quantify the impact of anatomical changes between pCT and MR acquisition. In addition, the MR image is also corrected for intensity variations across the image (cf. section 4.4). The necessity of this correction is investigated and its impact quantified. All sCT images are compared to the pCT and the respective BD-CTs.

The workflow is summarized in the flowchart in figure 3.1.



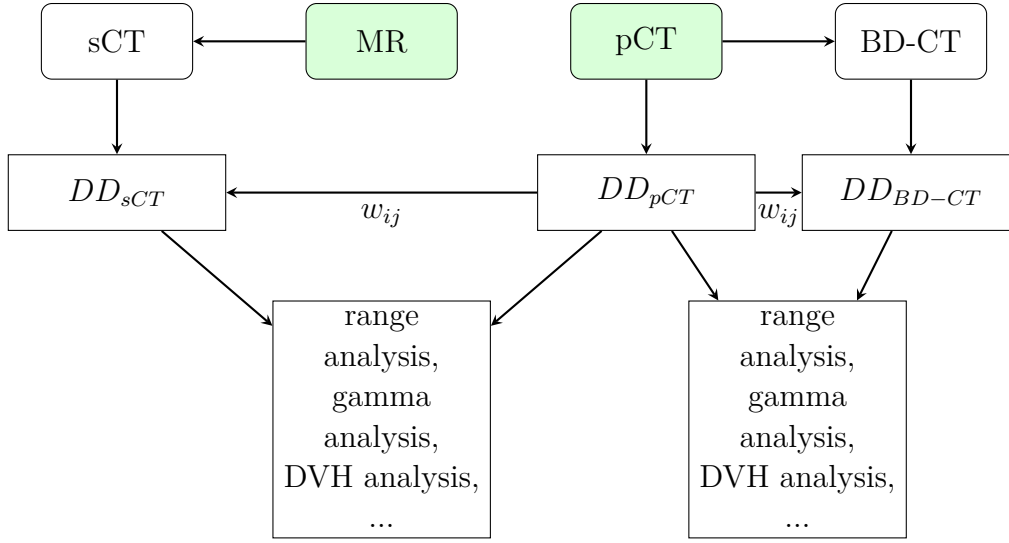


Figure 3.1.: Processing of the clinical data (green boxes) used as inputs for the workflow.

The workflow was implemented such that a modular framework for BD-CT or sCT generation and analysis was generated. This enables easy adaption to various datasets and different treatment modalities. In the future, it also allows to repeat the investigation, when individual modules have been exchanged. An exchange of modules might be necessary, if for example the module that generates sCT images from MR images is adapted to a new algorithm or new input images.

As a first step, the analysis was performed for photon therapy using a dataset acquired during the MR guidance study performed at the German Cancer Research Center. This step serves as a reference of plausibility of results, as much literature is published for photon therapy.

The analysis of BD-CTs for proton therapy was performed successively on two datasets: the Heidelberg Ion-Beam Therapy Center (HIT) dataset and the Gold Atlas dataset. The HIT dataset contains CT images and treatment plans from patients irradiated at the HIT, however no MR images. The Gold Atlas dataset is a publicly available data set (Nyholm et al., 2018) containing CT and MR images, however no treatment plans. The latter have been created manually on the CT images to repeat the BD-CT study also on this dataset. The results of the BD-CTs from the HIT dataset thus serve as benchmark of plausibility of the treatment plans. The MR images contained in the Gold Atlas dataset are the basis for the sCT generation.

The re-calculation of the nominal treatment plan from the pCT on first the BD-CT and second the sCT images makes it partly possible to quantify the errors of different sources. The BD-CT has the exact same anatomical configuration as the pCT, thus the difference in dose distribution is solely due to the BD overwrites. Assuming an MR image with little artefacts and the same anatomy, only the algorithm to detect

the tissue classes for BD overwrite introduces errors in addition to the pre-quantified errors of the BD values. The one or the other can then be refined upon demand.

## 4. Material and Methods

### 4.1. Patient data

Three different datasets with anonymised patient data were used in this study for the three steps of investigation described in the chapter above. The first one focused on the applicability of bulk density computed tomographies (BD-CTs) for intensity-modulated radiotherapy (IMRT), therefore only computed tomography (CT) and clinical IMRT plans were needed, which was available in the so-called "magnetic resonance (MR) guidance" dataset. For the second step, the analysis of BD-CTs for proton therapy, ideally clinical treatment plans are used to avoid a bias of results due to unrealistic treatment plans. Thus, clinical data from the Heidelberg Ion-Beam Therapy Center (HIT) was used. For the third step, the generation of synthetic computed tomography (sCT) images from MR, MR images are a crucial requirement of the dataset. For this reason, the open-source Gold Atlas dataset was used.

#### MR guidance study

The first part of this thesis investigated the impact of BD-CTs in IMRT. The data used for this part was part of the so-called "MR guidance" study (Bostel et al., 2014, 2018) conducted at the German Cancer Research Center. In this study, patients received regular MR imaging before the treatment and were then transported to the treatment site with a shuttle to avoid anatomical changes due to re-positioning on the treatment couch. From this dataset, six pelvic patients were selected for the analysis. Each dataset included a planning computed tomography (pCT), contoured target and organ at risk (OAR) structures, and a clinical IMRT plan. Details for the indication, prescribed dose and fractionation scheme can be found in table 4.1.

A common set of delineated structures is available for the six patients, on which structure-specific analysis is performed: the planning target volume (PTV), and the bowel, urinary bladder and spinal cord as OARs. Other desirable structures for analysis, such as the clinical target volume (CTV), are not available for all patients.

Table 4.1.: Indication, fractionation scheme (number of fractions times prescribed dose per fraction) and number of beams for IMRT treatment for the patients from the MR guidance study.

Patient	Indication	Fractionation scheme	Number of beams
MRG1	Rectum carcinoma	5 x 5 Gy	7
MRG2	Rectum carcinoma	28 x 1.8 Gy	7
MRG3	Rectum carcinoma	28 x 1.8 Gy	8
MRG4	Prostate carcinoma	34 x 2.25 Gy	12
MRG5	Cervix carcinoma	28 x 1.8 Gy	9
MRG6	Anal carcinoma	28 x 1.8 Gy	9

## HIT data

For the BD-CT analysis for proton therapy, four patients treated for prostate cancer and one patient treated for cervical cancer at the HIT were used. Each patient dataset included a pCT, contoured target and OAR structures, and a clinical intensity-modulated proton therapy (IMPT) plan. The prostate cancer patients were treated with two opposing beams (cf. figure 4.1) and a 26 x 2.7Gy(RBE) fractionation scheme; the cervical cancer patient was treated with V-shaped beam configuration (cf. figure 4.2) with beam angles 160° and 200° (with 0° defined as 12 o'clock in the figures, going clockwise) and a fractionation scheme of 28 x 1.8Gy(RBE). An overview of the specifications of each patient can be found in table 4.2.

Table 4.2.: Indication, fractionation scheme (number of fraction times prescribed dose per fraction) and beam angles for the IMPT treatment for the patients treated at HIT.

Patient	Indication	Fractionation scheme	beam angles
HIT1	Prostate carcinoma	26 x 2.7 Gy(RBE)	90°/270°
HIT2	Cervix carcinoma	28 x 1.8 Gy(RBE)	160°/200°
HIT3	Prostate carcinoma	26 x 2.7 Gy(RBE)	90°/270°
HIT4	Prostate carcinoma	26 x 2.7 Gy(RBE)	90°/270°
HIT5	Prostate carcinoma	26 x 2.7 Gy(RBE)	90°/270°

## Gold Atlas

The Gold Atlas project (Nyholm et al., 2018) provides publicly available datasets of 19 patients (18 prostate carcinoma, 1 rectal carcinoma) in total, acquired at three sites in Sweden. In this study, patients from the Gold Atlas dataset are denoted as GA1.x (x = 1-8) for site 1, GA2.x (x = 1-7) for site 2, and GA3.x (x = 1-4) for

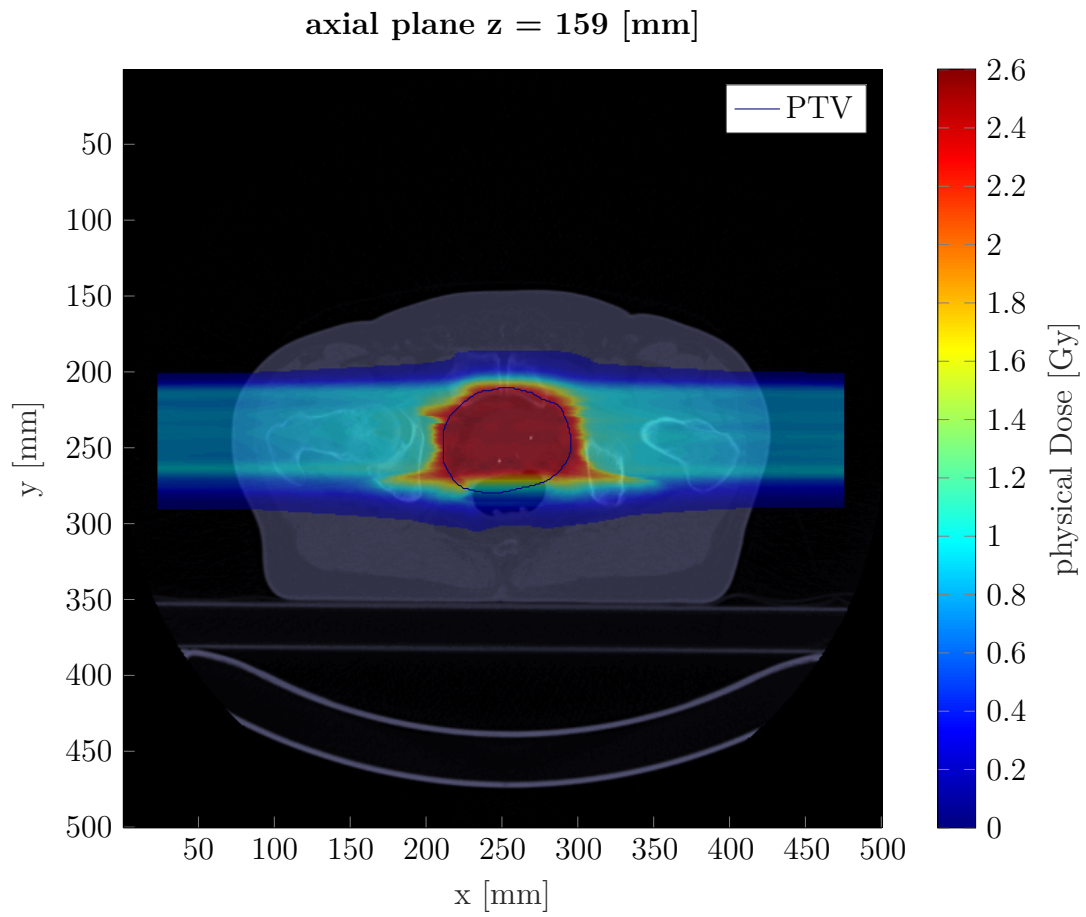


Figure 4.1.: Dose distribution in color-overlay on CT image of a prostate cancer patient from the HIT dataset.

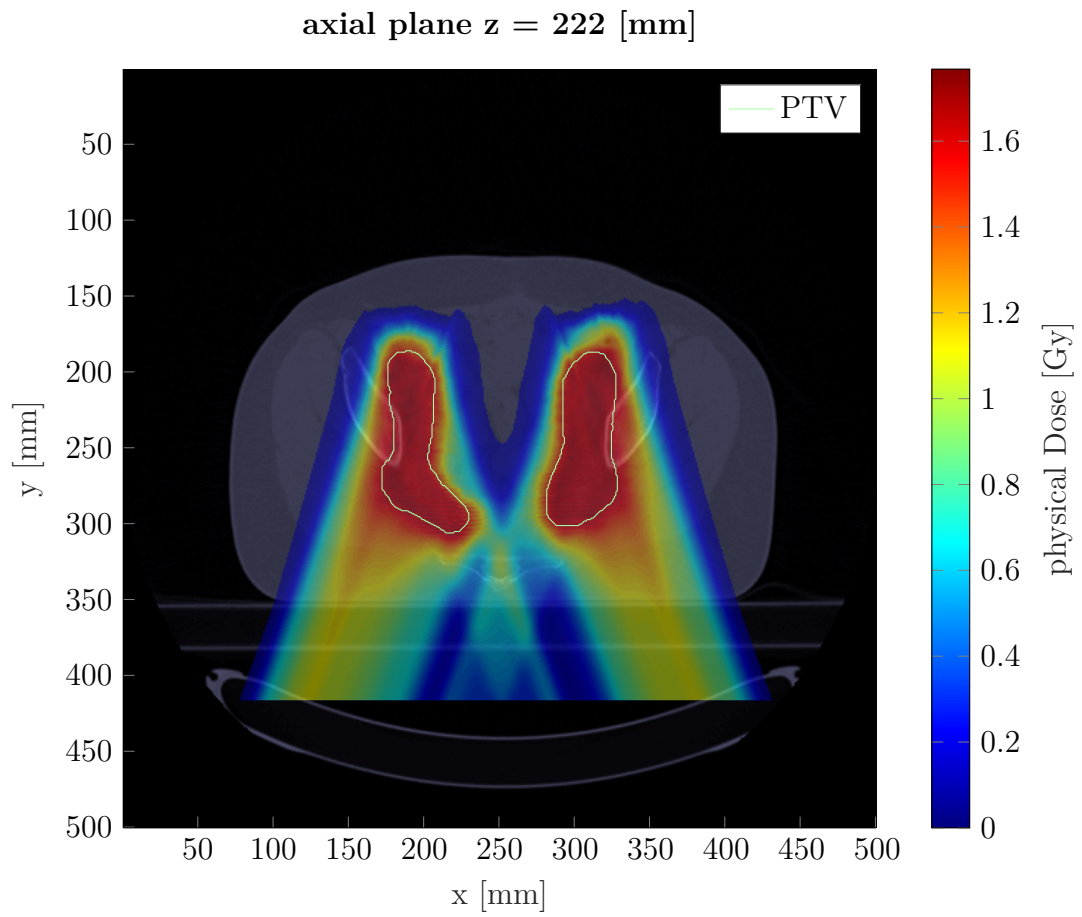


Figure 4.2.: Dose distribution in color-overlay on CT image of a cervical cancer patient from the HIT dataset.

site 3. The datasets contain T1- and T2-weighted MR images, CT images, a rigid registration between CT and MR, a deformably registered CT to MR image, and contours of organs of interest. The specifications (B0 field strength,  $TE$  and  $TR$  times, etc.) of the MR images vary slightly from site to site, details can be found in Nyholm et al. (2018). The most important difference between the three sites is that at site 1 and 3, the B0-field strength of the scanner is 3T, whereas at site 2 it is 1.5T. In addition, the T2-weighted MR images hold multi-observer delineations from five different clinicians and consensus delineations of organs; in the CT image, only the body contour and the bladder are delineated. The original delineations from the T2-weighted MR images were copied to the T1-weighted MR images for the structures used for the sCT images derived from the original MR images. For each site, a parameter file for the deformable registration in *elastix* (Klein et al. (2010), cf. section 4.3) is available. This was used to deformably register the T2-weighted MR (and its delineations) to the CT. This was a necessary pre-processing step for optimizing dose distributions on the CT. From the deformed delineations, the prostate was used as the CTV. An isotropic dilation of the CTV by 5mm was performed to create the PTV. The overlap of the 1mm isotropic dilation of the PTV and the rectum was defined as the rectal anterior wall, the remaining rectal volume as rectal posterior wall. The body contour and bladder contour of the CT were used instead of the deformed ones for higher agreement. An example of these target and OAR definitions is shown in figure 4.3.

## 4.2. Treatment planning tool: matRad

matRad is an open-source treatment planning tool (Wieser et al. (2017), <https://github.com/e0404/matRad>) for educational and research purposes developed at the German Cancer Research Center. In this thesis, it is used for plan optimization and dose re-calculation, as well as a platform for additional developments. Scripts and workflows in this thesis, for example the generation of synthetic CT images, are integrated into the matRad environment.

matRad is written in matlab (The Mathworks, Inc., <https://mathworks.com>) and offers both a graphical user interface and command line interaction for more efficient work with larger data sets or standardized workflows. It supports intensity-modulated photon, as well as scanned proton and carbon ion beam radiotherapy.

The workflow of a general matRad calculation can be explained on the basis of the matRad variables. At first, an image has to be imported from the Digital Imaging and Communications in Medicine (dicom) data<sup>a</sup>, which is the standard format for storing medical image data. Image data, dose distributions, organ delineations, dose prescription, and many more relevant information can be stored in the dicom

---

<sup>a</sup><https://www.dicomstandard.org/>

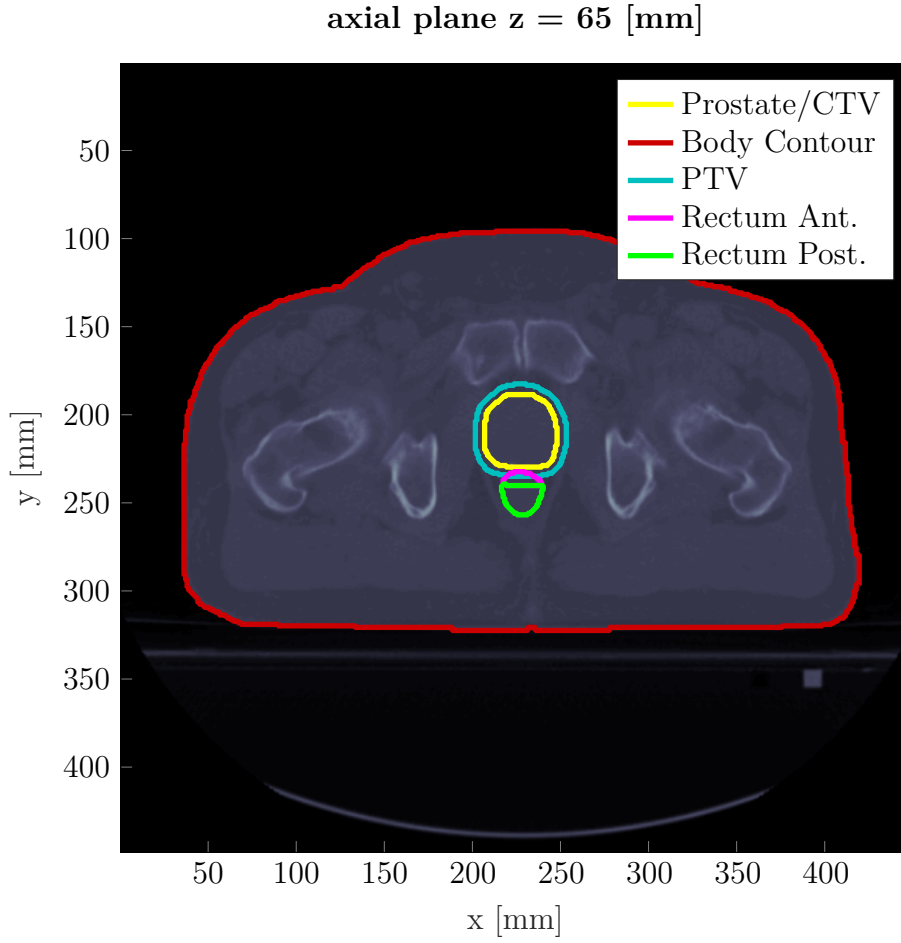


Figure 4.3.: Example slice from one patient of the Gold Atlas dataset (Nyholm et al., 2018), illustrating the definition of target (CTV, PTV) and organ at risk (rectal anterior and posterior wall) delineations.

format. From the dicom file a "ct" structure is created, which contains the image cube and meta data (e.g. resolution, cube dimension, etc.). Within matlab, this can also be manipulated, for example when replacing the CT with a sCT. In the "cst" structure, the volumes of interest are defined. Next, the plan parameters can be imported from dicom file or manually set to the "pln" structure, including amongst others beam and gantry angles, and particle type. Based on these three inputs, a steering structure "stf" is generated, which contains detailed information about each beam (consisting of multiple pencil-beams) used in the plan, such as position, energy, beam source points and target points of the pencil-beams. This information can be read from a dicom file if available, or it can be calculated based on ct, cst and pln information. matRad utilizes an inverse planning approach, therefore the dose influence matrix "dij" or  $D_{ij}$ , which describes the influence of a pencil-beam "j" on a finite volumetric element (i.e. on a voxel) "i" for all pencil-beams of the plan and all voxels in the image, is calculated. The dose in voxel "i" can then be calculated as in equation 2.13. The size of all pencil-beam fluences within  $w_j$  has to be optimized in



a next step using an iterative optimization algorithm, to achieve a dose distribution that satisfies all prescriptions to the target volume and constraints to OARs defined by the clinician. After optimization, final dose distributions and the fluences  $w_j$  are saved in the "resultGUI" structure.  $w_j$  can be used for re-calculating an optimized dose distribution on a different CT image. Therefore,  $D_{ij}$  has to be calculated for the new image, the "re-calculation" is then a single matrix multiplication of the new  $D_{ij}$  from the different CT image and the old  $w_j$  optimized on the original image. Furthermore, matRad contains many analysis scripts, for example to calculate dose volume histogram (DVH) curves or gamma analysis (both cf. section 4.7). Since many processes have been automated or adapted in this thesis, only the command line interaction was used and workflow scripts based on matRad functions were created.

### 4.3. Rigid and Deformable Registrations

Independent from the chosen method to create synthetic CT images, image registration is always necessary. Image registration is the process, that brings two images (from different time points, of different modalities, etc.) to best spatial agreement. Rigid registrations, i.e. translation and rotation of images, are necessary to overlay CT and MR images from different, machine-dependent spaces. Often, rigid registrations alone are not sufficient, since they can not describe anatomical changes as they can occur with organ motion or weight loss. For this purpose, deformable registrations have been developed, which aims at matching anatomical structure over each other by locally deforming the image. Any registration consists of three main parts (Oh and Kim, 2017): an optimization algorithm, a similarity measure and a transformation model. The optimization algorithm tries to maximize the similarity measure between the fixed image and the moving image (i.e. the one whose structure will be changed to match the fixed image), which has been deformed according to the transformation model.

Generally, the transformation model is not invertible as it is not a bijective mapping. The most common similarity measures are the sum of (voxel-wise) squared differences for mono-modal registration, i.e. for registrations between two images of the same modality such as CT to CT, and mutual information for multi-modal registration, i.e. for images of two different modalities, such as CT to MR.

For the MR and CT data sets of the Gold Atlas (cf. section 4.1), a rigid transformation is available with the data, which was applied in this study as is. Additionally, the Gold Atlas also provides CT images, which have been deformably registered to the MR, and the corresponding parameter files for the registration. In this thesis, however, a deformable registration of the MR to the CT was needed (i.e. fixed and registered image switched), therefore the parameter files were used to generate new, registered MR images. The workflow of the Gold Atlas registration was followed

as described in Nyholm et al. (2018) (described in full detail in appendix A), only the registered and fixed images were switched. Briefly, the urinary bladder contours from the MR and CT images are added in binary format to the registration input, as well as the CT body contour to mask the region of interest. The registration is performed on three levels of resolution for more robustness, from coarse to original resolution. The registration algorithm uses a BSpline approach (i.e. the deformation can be described with continuous, piece-wise polynomial functions), which is independent from complete segmentation of the input images. As similarity measures, mutual information is applied for the CT/MR comparison, whereas for the bladder binaries, mean squared differences are calculated. The output of the algorithm is a transform file, specifying the BSpline-parameters of the deformation, which is then applied to the MR image and all binary segmentations of the MR. The registration was performed using the *elastix* software package<sup>b</sup> (Klein et al., 2010).

## 4.4. Bias Field Correction

One MR artefact mentioned in section 2.1, which does not affect the anatomical information of the image and therefore generally is not corrected for, is the low frequency non-uniformity of the image. It originates from an inhomogeneity of the fields produced by the radio-frequency (RF) coils or inhomogeneity of the  $\vec{B}_0$ -field of the scanner, therefore the expression of the artefact is proportional to the strength of the  $\vec{B}_0$ -field. It does, however, affect quantitative use of the image, since grey values alone are not sufficient any more for same tissue identification. The most common algorithm to correct this artefact is the N4 bias field correction (BFC) algorithm introduced by Tustison et al. (2010). It is implemented in the open-source medical image processing software *Slicer*<sup>c</sup>, which was applied in this thesis for this purpose. The corrective bias field, or the corrected image, is obtained in an iterative process. Based on the assumption that a (noise-free) measured image  $v(\vec{x})$  is a product of the uncorrected images  $u(\vec{x})$  and the smooth, slowly varying bias field  $f(\vec{x})$ :

$$v(\vec{x}) = u(\vec{x})f(\vec{x}). \quad (4.1)$$

Using the notation  $\hat{v} = \log(v)$ , this becomes

$$\hat{v}(\vec{x}) = \hat{u}(\vec{x}) + \hat{f}(\vec{x}). \quad (4.2)$$

The solution for the corrected image  $\hat{u}(\vec{x})$  in the  $n^{\text{th}}$  iteration of the BFC algorithm is calculated as

$$\begin{aligned} \hat{u}^n &= \hat{u}^{n-1} - \hat{f}_r^n \\ &= \hat{u}^{n-1} - S * \{\hat{u}^{n-1} - E[\hat{u}|\hat{u}^{n-1}]\}, \end{aligned}$$

---

<sup>b</sup><http://elastix.isi.uu.nl/>

<sup>c</sup><https://www.slicer.org/>

where  $\hat{f}_r^n$  is the residual bias field (i.e. the bias field needed to correct the updated image  $\hat{u}^{n-1}$  from the previous iteration),  $S * \{\cdot\}$  denotes a B-spline approximator, which satisfies the requirement of a smooth and slowly varying bias field, and  $E[\hat{u}|\hat{u}^{n-1}]$  is the expected value of  $\hat{u}$  given the current estimate of the corrected image ( $\hat{u}^{n-1}$ ). The total bias field is then the sum of the residual bias field of each iteration. An example for an MR with bias field, a corrective bias field and the corrected MR is given for one patient of the Gold Atlas data set in figure 4.4.

Bias field correction was applied in this thesis to all used MR images, to investigate the influence of this image pre-processing step on the quality of the sCT images.



Figure 4.4.: (a) T1-weighted MR image from the Gold Atlas dataset. (b) Bias field of MR image. (c) Bias field corrected MR image.

## 4.5. Own developments for this thesis

### 4.5.1. Bulk Density Computed Tomographies

The term "synthetic CT" implies that it is a general goal to produce images, which look indistinguishable from an original CT image. This fact is strongly reflected by the metrics used for evaluation: one of the most commonly reported metrics is the mean absolute error (MAE) between the CT and the sCT (Edmund and Nyholm, 2017):

$$MAE = \frac{1}{N} \sum_{i=1}^N |sCT(i) - CT(i)|, \quad (4.3)$$

with  $N$  the total number of voxels in the images, and  $sCT(i)$  and  $CT(i)$  the HU values at voxel  $i$ , respectively. It is, however, not clear whether it is necessary for dose calculation to reproduce over 4000 different grey values for clinically comparable dose distributions.

The motivation for this question can be explained with the help of a histogram of rED numbers, i.e. the grey value of the voxels in the CT converted using a Hounsfield look-up table (HLUT), as seen in figure 4.5 for the body contour of a

pelvic cancer patient. Two peaks are clearly dominant: the one around  $rED \sim 0.95$  represents fat, the one around  $rED \sim 1.05$  represents soft tissues.

This seems reasonable considering that the shape is determined by two factors: the distribution of true CT numbers of the tissue and by artefacts, most dominantly noise, which are Gaussian shaped (Diwakar and Kumar, 2018). For proton therapy, the sum of the stopping power ratio (SPR) values (for photon therapy the relative electron density (rED) values, respectively) inside the voxels along the path of the beam determine the particle range for a fixed energy. If these SPR and rED values vary symmetrically around a mean, than the sum of them is equal to the summation of the mean values. This assumption can be made for Gaussian distributions, as they are symmetrical.

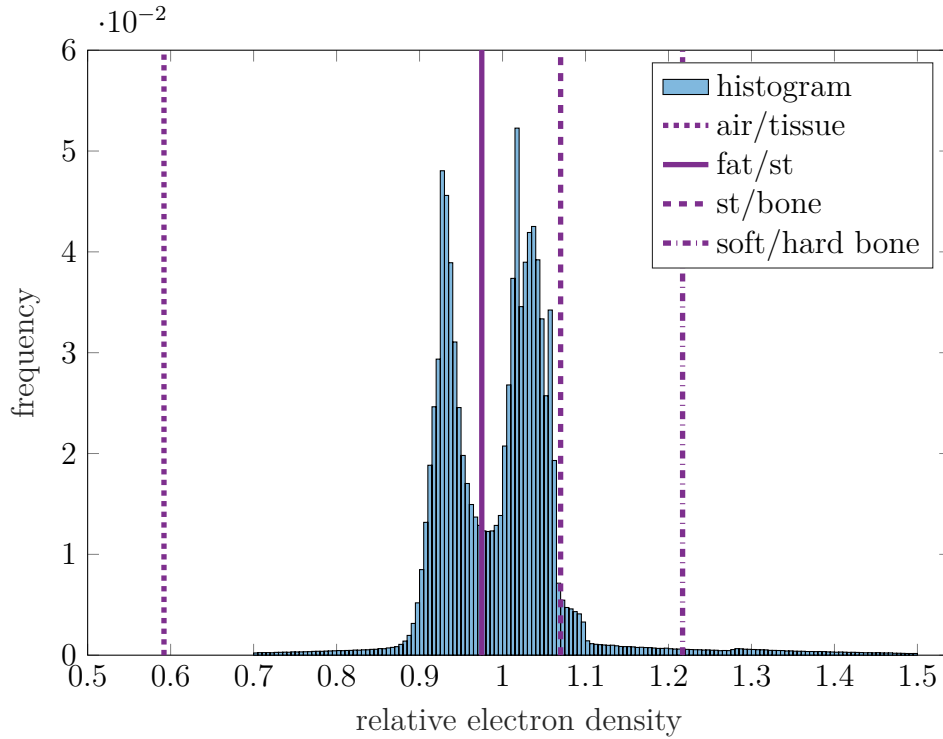


Figure 4.5.: Histogram of the rED numbers within the body contour of one patient from the MR guidance dataset. The lines indicate the thresholds for used for tissue segmentation. Air is segmented below the dotted line, between the dotted and the solid line is fat, between the solid and the dashed line is non-fatty soft tissue (st), above that bone which is further slit at the dotted-dashed line in soft and hard bone.

To address the question of the necessary number of bulk densities (BDs), the number of BDs in the pCT was varied starting from the maximum reduction comprising only one class. As measure of quality, the original dose plan was re-calculated on the CT containing the BD classes (BD-CTs) and compared to the original.

The pCT was segmented into tissue classes based on its grey values. The segmen-

tation into five classes was executed in the following steps:

1. Bone was pre-segmented above  $\text{SPR} = 1.089$  and  $\text{rED} = 1.069$ , respectively, producing a skeleton of the bone segment. Since this does not generally cover soft bone, which has lower grey values due to its high fat content, the resulting bone contours were filled using morphological operations:
  - a) Small gaps of maximum 3 pixels (which might for example occur where the cortical bone is very thin) were closed, so the segmentation skeleton is fully connected.
  - b) The skeleton was filled to remove the hollow parts
  - c) Single pixels or small clusters of pixels were removed. Those single pixels or clusters might have been included to the segment during the application of the  $\text{SPR} = 1.089$  and  $\text{rED} = 1.069$  threshold, for example where small metallic implants have been used as position markers.
  - d) Within the complete bone segmentation, voxel above  $\text{SPR} = 1.4$  and  $\text{rED} = 1.217$ , respectively, were classified as cortical bone, the remaining part was classified as spongy bone.
2.  $\text{SPR} = \text{rED} = 0.6$  inside the body contour was assigned air.
3. Based on the histogram, for each patient the minimum position between the fat and the soft tissue peak was determined as the optimum splitting point for those tissue classes (more details how this point was determined can be found in the following section on BD value determination).
4. Tissue below the splitting point and above the air threshold was assigned fat.
5. Tissue above the splitting point that is not bone was assigned soft tissue.

For the BD-CTs with different granularity, the segments were assigned different BD values. For the investigation for conventional IMRT with photons, four different BD-CTs were created with one to four BDs: (1) one water-equivalent, (2) one single BD, (3) one with two BDs for fat and soft tissue, and (4) one with four BDs for fat, soft tissue, air and bone. With knowledge of the results from BD-CTs for the photon plans, the BD-CTs were adapted subsequently for the proton plans. Here, BD-CTs with two BDs or less were omitted. One BD-CT with four BDs (air, fat, soft tissue, bone) and one with five BDs (air, fat, soft tissue, spongy bone, cortical bone) were created. In addition, BD-CTs with original content (either bone or soft tissue were left as in the pCT) were created, to investigate the influence of the different BD groups. An overview over all BD-CTs can be found in table 4.3.

The BD numbers to overwrite the different segmentations in the photon study were chosen to be patient-specific  $\text{rED}$  values for fat and soft tissue, and fixed values for air and bone. For the proton study, patient-specific  $\text{SPR}$  values were used for each BD. From a clinical point of view, the CT image as a trustworthy reference will not

Table 4.3.: Overview over the different bulk density (BD)-CTs created for intensity modulated radiotherapy with photons (top) and proton therapy (proton). For each BD-CT (rows), the content of the segmented classes (columns) is given: "BD" refers to a bulk density overwrite value for the whole segment, "pCT" refers to the original content from the pCT.

	air segment	fat segment	soft tissue segment	spongy bone segment	cortical bone segment
photon					
WE	BD				
1t	BD				
2t	BD		BD		
4t	BD	BD	BD	BD	
proton					
WE	BD				
ST	pCT	BD	BD	pCT	pCT
B1	pCT	pCT	pCT	BD	
B2	pCT	pCT	pCT	BD	BD
STB1	BD	BD	BD	BD	
STB2	BD	BD	BD	BD	BD

be abandoned in the near future. Thus, patient-specific BD values will be available in all cases.

#### 4.5.2. Determination of bulk density values

The BD overwrite values for the BD-CT images (explained in the previous section) have to be determined from the image. In this thesis, a fitting approach of the rED and SPR histograms for photon and proton therapy, respectively, has been chosen. After conversion of the pCT image using the clinical HLUT as explained in section 2.3.1 and 2.3.2, the rED and SPR values were only analysed within the body contour delineated by the physicians. The two dominant peaks representing fat and soft tissue (st), as can be seen in the rED histogram in figure 4.6 for a patient treated with photons, are fitted with a sum of two Gaussian functions (orange line), a so-called two-termed Gaussian function:

$$G(x) = c1 \cdot \exp\left(-\frac{(x - \mu_{fat})^2}{2\sigma_{fat}^2}\right) + c2 \cdot \exp\left(-\frac{(x - \mu_{st})^2}{2\sigma_{st}^2}\right) \quad (4.4)$$

The mean values  $\mu_{fat}$  and  $\mu_{st}$  correspond to the peak positions (dotted and dashed purple lines). Potential noise or artefacts in the image are neglected this way. To separate the two tissue classes fat and soft tissue, the minimum point between the two Gaussian functions was determined as a threshold (solid purple line).

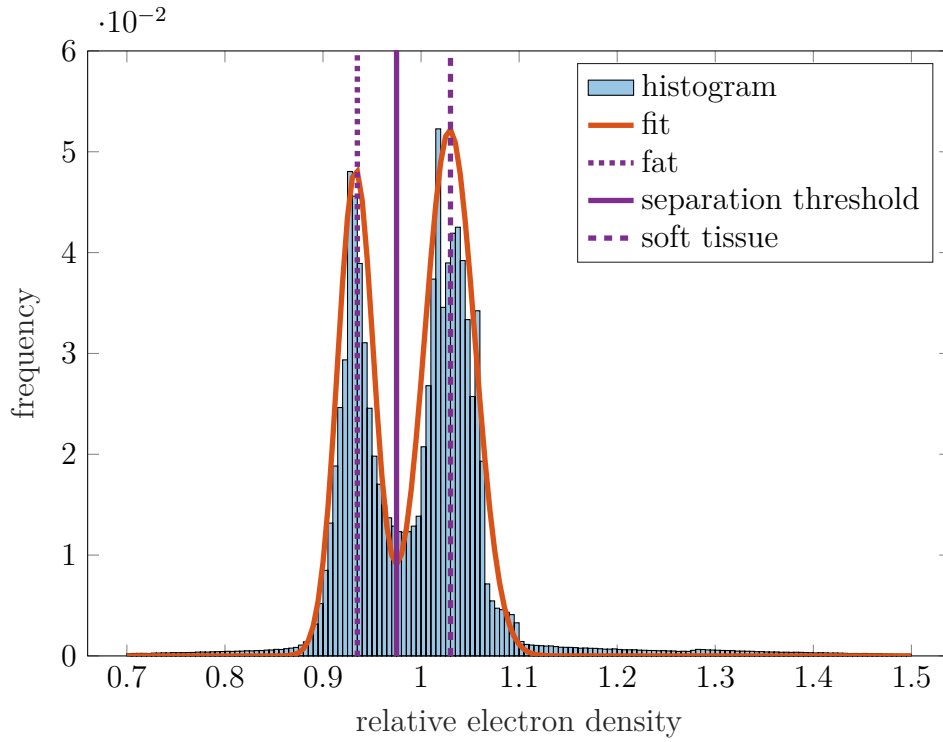


Figure 4.6.: Example of relative electron density histogram (blue) and fit of the two peaks using a two-term Gaussian function (orange). The peak positions, i.e. the  $\mu$  fit parameter of the Gaussian functions, for fat (dotted line) and soft tissue (dashed line), and the threshold used for grey-value based separation of the two tissue classes (solid line) are shown as vertical lines at respective relative electron density values.

The usage of BD values as overwrite values for tissue classes was one of the earliest investigated methods to create sCT images. How these BD values are derived changes from study to study. Most commonly, mean values (Lee et al., 2003; Doerner et al., 2015) of tissue classes derived from the CT image, or reference values from International Commission on Radiation Units & Measurements (ICRU) report published by White et al. (2016) are applied (Jonsson et al., 2010; Korsholm et al., 2014). In this study, the peak position  $\mu_{fat}$  and  $\mu_{st}$  are used as BD values for the fat and the non-fatty soft tissue, respectively. In comparison to the usage of mean values or reference values, the fit approach has certain advantages. First, it is independent from pre-segmentation of the individual classes, based on which average values can be determined. Secondly, it is more robust towards artefacts or an "overpronunciation" of individual rED or SPR bins, that could distort the mean value. The usage of the ICRU report is limited in the sense that only very specific types, e.g. the femur as surrogate of bone, are listed in that report, which do not reflect the entirety of rEDs or SPRs.

### 4.5.3. Generation of synthetic CTs

The sCTs derived from MR images were generated based on the T1-weighted MR images from the Gold Atlas dataset (cf. section 4.1). The sCT generation tried to mimic the BD-CT containing five tissue classes (i.e. air, fat, non-fatty soft tissue, spongy bone, cortical bone). At first, the bones were segmented via an intensity threshold (below the red vertical line in figure 4.7), which was defined as the signal intensity on the first rising slope of the histogram at 1/3 of the maximum height of the corresponding peak. Due to the ambiguity of air and bone intensity values in the MR, these two classes were separated using a bone probability mask. The mask was generated by adding the 19 segmented bones masks from the CT image of the Gold Atlas patients and normalizing the result. Thus, the voxels contains values between 0 and 1, giving the cohort probability of bone in every voxel (0 - no patient had bone in this voxel, 1 - all 19 patients had bone in this voxel). If a voxel was segmented based on the threshold and the probability was larger than 0.5 (blue area in figure 4.8), it was classified as cortical bone, otherwise as air.

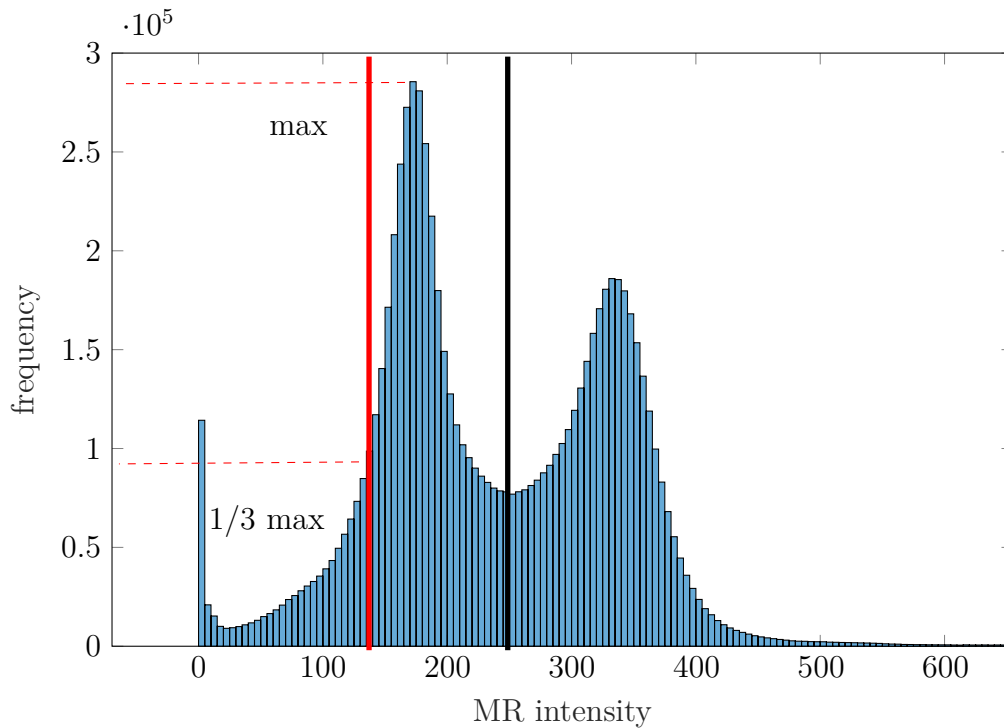


Figure 4.7.: Histogram of a bias field corrected T1-weighted MR image from the Gold Atlas dataset. The red vertical line marks the initial threshold where air/cortical bone is separated from the remaining tissue; the black line marks the threshold that is used to separate muscle from fat/spongy bone.



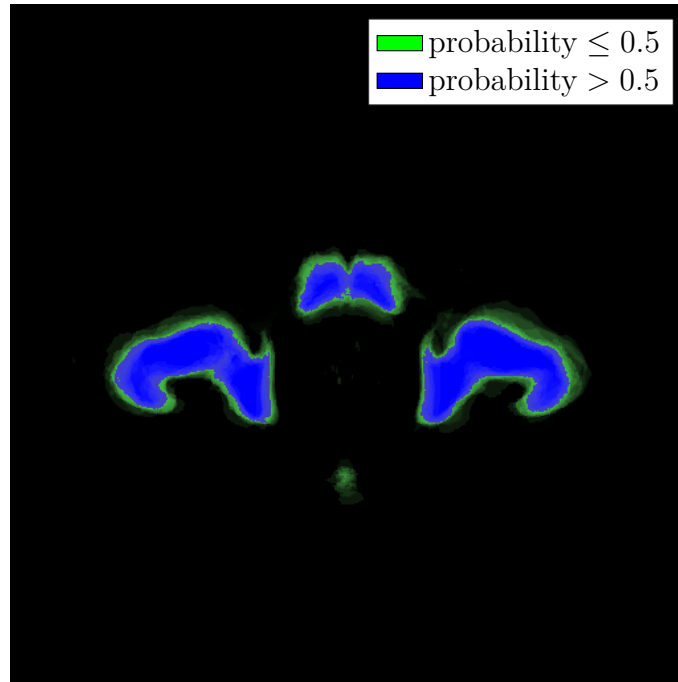


Figure 4.8.: One slice from the probability mask for bone. The area where the probability is  $> 0.5$  is shown in blue,  $\leq 0.5$  in green.

Since spongy bone contains a lot of fat, the signal intensity is almost indistinguishable from subcutaneous or visceral fat. Therefore, a second threshold was applied to separate potential fatty tissue from other soft tissue, bone and air (the black line in figure 4.7), which was defined as the minimum position between the two peaks. Tissue above this threshold is fatty, including spongy bone, whereas tissue below, i.e. between the red and the black vertical lines, is mostly muscle and other non-fatty soft tissue. The bone probability mask was then used again to separate spongy bone (probability  $> 0.5$ ) from other fat (probability  $\leq 0.5$ ).

The five tissue classes segmented on the clinical MR were subsequently overwritten with BD values, derived patient-specifically from the SPR/rED histogram. The BD values correspond to the respective values in the BD-CT. The same procedure was applied to the deformably registered MR (cf. section 4.3), to minimize anatomical differences between the two images. To minimize the impact of the bias field (cf. section 4.4), the same procedure was also applied to the bias field corrected MR. Finally, a sCT was generated based on a deformably registered and bias field corrected MR image. Thus, in total, four sCT images were generated:

- sCT<sub>O</sub> using the original MR
- sCT<sub>R</sub> using the registered MR
- sCT<sub>B</sub> using the bias field corrected MR
- sCT<sub>BR</sub> using the registered and bias field corrected MR

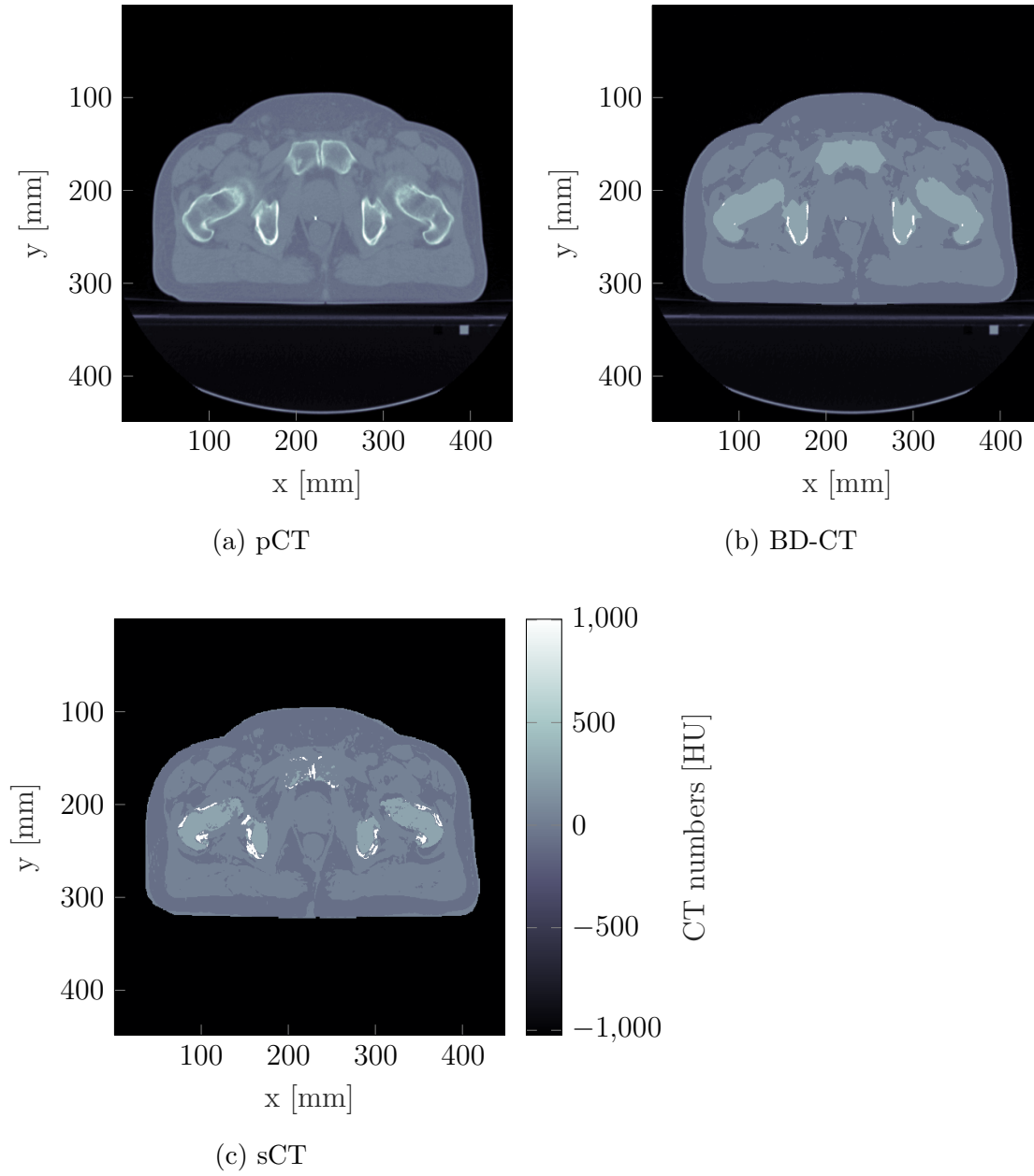


Figure 4.9.: Examples of (a) a planning CT (pCT), (b) a bulk density CT (BD-CT), and (c) a synthetic CT (sCT) image for the pelvis for patient GA1.1 of the Gold Atlas dataset.

## 4.6. Dose calculation and re-calculation

In an ideal, MR-based treatment planning workflow, the dose optimization and calculation is performed on the sCT image. Taking a real CT, i.e. the pCT, as the gold standard, the error in the dose distribution (DD) caused by the sCT image can be analysed by a re-calculation of the DD optimized on the sCT and re-calculated on the pCT.

In the first part of the thesis, this approach was implemented for the BD-CT study for photon therapy. For each BD-CT, a DD was optimized using clinical dose prescriptions and constraints to imitate the clinical plan from the pCT. The beam weights from the clinical plan was used for beam weight initialisation during each optimization, to achieve as similar DDs as possible. A 1%/1mm gamma analysis (explained in section 4.7) was performed to compare the DD from the pCT ( $DD_{pCT}$ ) and the DD from the BD-CTs. All comparisons yielded a 100 % gamma pass rates (GPRs) in the PTV and inside the volume that receives more than 2 % of prescribed dose, thus the DD from the BD-CTs are considered equal to the  $DD_{pCT}$ . In a next step, these DD are re-calculated on the pCT using the beam weights  $w_j$ . Depending on which BD-CT the DD was optimized on, the re-calculated DD are labelled  $DD_{BD-CT}$ . For example: the DD optimized on BD-CT<sub>WE</sub> is called  $DD_{WE}$  after re-calculation on the pCT.

The same approach is not feasible for proton therapy, for efficiency reasons. The optimization of a proton treatment plan is more time-consuming, as the plan has many more degrees of freedom. For this reason, the DD of the pCT was re-calculated on the BD-CTs. The resulting DD is also labelled  $DD_{BD-CT}$ , although it should be noted that the effect of the BD-CT on the dose difference is now reversed: a decrease in radiological depth in the BD-CT would lead to an undershooting of the proton beam when re-calculating the BD-CT-based plan on the pCT, whereas for the re-calculation of the pCT-based plan on the BD-CT, the proton beam would overshoot.

For the HIT dataset, the clinical treatment plan was re-calculated within matRad and the resulting DD was used as the reference  $DD_{pCT}$ . It was then re-calculated on the BD-CTs. The reference DD for the Gold Atlas dataset had to be optimized manually, as no clinical data was published. The optimization was performed using prescriptions and constraints as they are applied at the HIT facility, to create realistic DDs.

The  $DD_{pCT}$  created for the Gold Atlas dataset have also been used for re-calculations in the sCT images. These DDs are labelled  $DD_O$ ,  $DD_R$ ,  $DD_B$  and  $DD_{BR}$ , respectively.

## 4.7. Analysis Metrics

Metrics in this thesis were selected to describe the DD after forward-calculation of the nominal dose on the BD-CTs/sCTs. DVHs were chosen, since they are widely used for dose prescription for targets and dose constraints for OARs. Thereby, differences of these criteria using BD-CTs/sCTs instead of the pCT can directly be detected. By nature, DVHs give no information on the location of observed dose differences, therefore the gamma analysis (cf. equation 4.5) is added, which offers location information, but also yields generalised results via GPR. As a last metric, the proton range analysis is included. A change in proton range, i.e. of the Bragg peak position, will shift the position of highest dose deposition, worst case it will be shifted to healthy tissue and miss the tumour. Proton range analysis creates two-dimensional (2D) range difference plots and mean values for comparison among larger data sets.

### Dose Volume Histograms and Dose Volume Metrics

Dose volume metrics (DVMs), derived from cumulative DVHs, are the most common measures to evaluate a given DD, but are also used for dose prescription to the target or normal tissue constraints. In such a histogram, the volume which receives at least a certain dose is plotted for selected structures. DVM then refers to dose in a specific volume, for example:  $D_{50}$  is the minimum dose, that 50 % of the volume receive. Certain DVMs are used as indicator for specific DDs: the  $D_2$  is an indicator of hotspots, and  $D_{98}$  is an indicator for the coverage of the structure with dose. The reference volume (corresponding to 100 %) can be the whole body, but also smaller volume of interests (VOIs), such as the target volume or OARs. An example for DVH curves of a target volume and an OAR (rectum in this case) is shown in figure 4.10 for one prostate cancer patient treated with protons at HIT.

### Gamma Analysis

Gamma analysis is a common tool for comparison of two DDs, not only for treatment plan comparisons, but also for example during quality assurance to compare a simulated plan to a measurement. It incorporates not only dosimetric differences, but rather a "distance to agreement", meaning that the dose is compared not only in the exact same position, but within a pre-defined radius around the reference position. The gamma index  $\gamma(\vec{r})$  was first introduced by Low et al. (1998) as

$$\gamma(\vec{r}) = \min\left\{\sqrt{\frac{|\vec{r}_c - \vec{r}|^2}{\Delta\hat{r}^2} + \frac{(d_c(\vec{r}_c) - d(\vec{r}))^2}{\Delta\hat{d}^2}}, \forall \vec{r}_c\right\}. \quad (4.5)$$

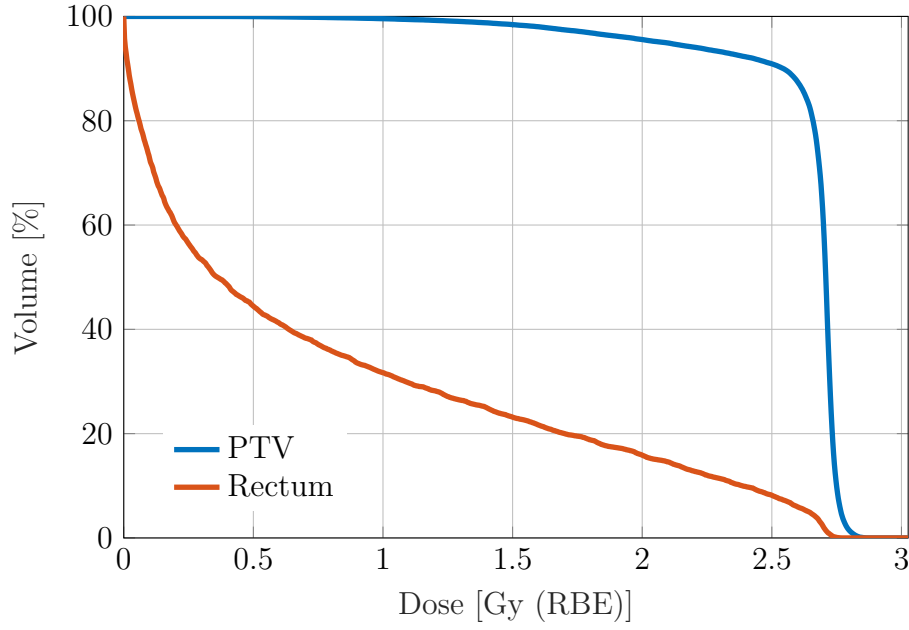


Figure 4.10.: Dose volume histograms for the planning target volume (PTV) and the rectum as an organ at risk for a patient treated for prostate cancer at HIT with a prescribed dose of 2.7Gy(RBE) per fraction.

$\vec{r}_c$  and  $d_c$  are the position and the dose to compare to, respectively,  $\Delta\hat{r}$  and  $\Delta\hat{d}$  are the pre-defined radius in geometrical and dosimetrical space, serving as acceptance threshold for the gamma pass criterion:

$$\begin{cases} \gamma(\vec{r}) \leq 1, & \text{passes} \\ \gamma(\vec{r}) > 1, & \text{fails} \end{cases} \quad (4.6)$$

As the gamma index is calculated for each voxel of the reference image, a three-dimensional (3D) gamma index cube is created, which can be analysed visually. An example of one slice of a gamma index cube is shown in figure 4.11 overlaid onto the corresponding slice of the pCT, which was used as the reference. The percentage of voxels within a certain volume that pass the criterion is called the GPR. Typical values for  $\Delta\hat{r}/\Delta\hat{d}$  are for example 2mm/2% or 1mm/5%. The latter criterion penalizes geometrical deviations stronger than dosimetrical deviations. The distribution of gamma values can also be plotted as 3D maps to visually correlate failure and position.

Generally, two types of gamma values can be distinguished: local and global values. The difference is the reference for  $\Delta\hat{d}$ , which could either be the dose at the investigated (local) position, or the (global) maximum dose. In this thesis, only the global gamma value is used. As a consequence, only areas with a dose larger than  $\Delta\hat{d}$  are included in the analysis.

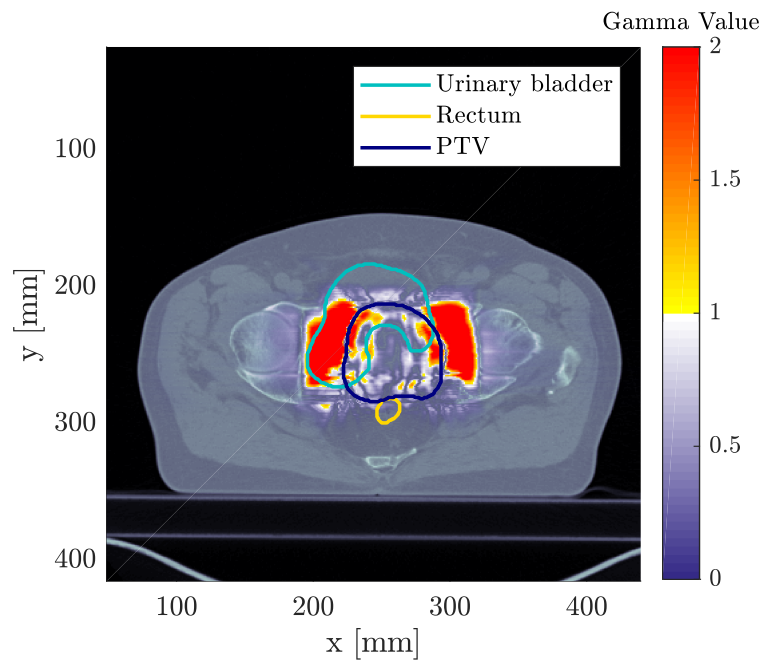


Figure 4.11.: Two-dimensional example of the gamma analysis performed using 2mm/2% criteria for BD-CT<sub>WE</sub> of patient HIT1 overlaid onto the corresponding CT slice.

## Proton Range Analysis

For the analysis of the range of the proton beam covering the target, the distal range for each bixel is extracted from the DD as the geometrical depth along the beam to the point where the dose distal to the dose maximum drops to 50 %. This point is dominated by the proton beam with the highest energy in that bixel. The range is then called  $R50$ . To quantify the changes of range compared to the pCT and the respective original DD, the differences between  $R50_{pCT}$  and  $R50_{BD-CT/sCT}$  are calculated bixel-wise. In addition, the mean and the standard deviation of the absolute difference of all pencil-beams of the plan are calculated to make the results more comparable among patients and BD-CTs/sCTs. In doing so, spatial information is lost and no correlation between anatomy and range shifts can be drawn. Therefore, 2D plots of the range shifts per beam in beam's-eye-view (BEV), are generated. An example of such a 2D range shift plot between pCT and a BD-CT with five BDs is shown in figure 4.12 for a prostate cancer patient treated at HIT.

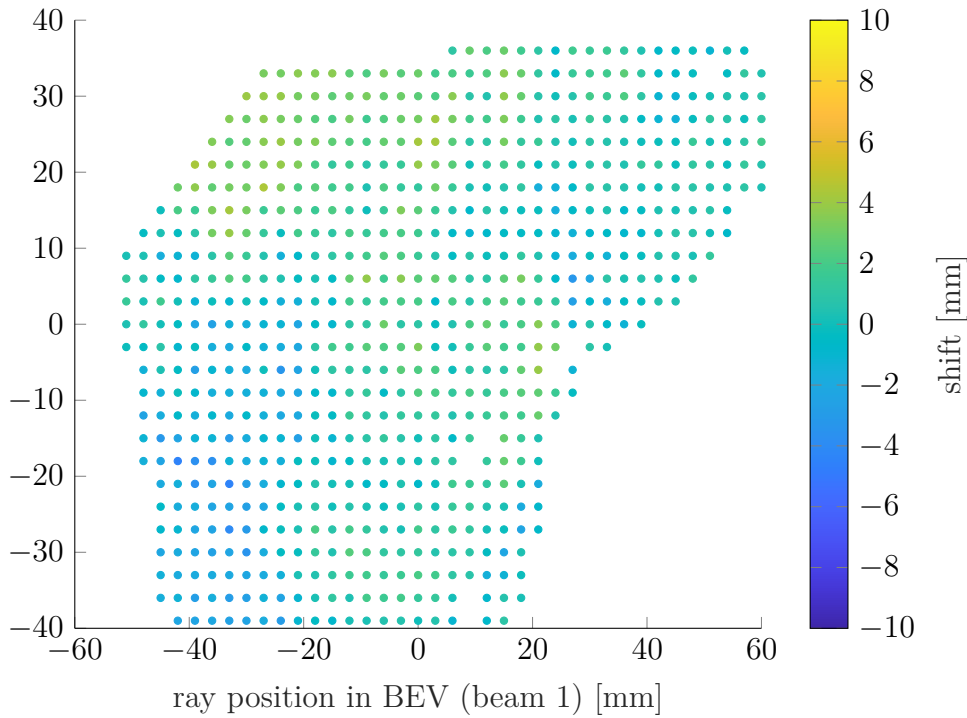


Figure 4.12.: 2D plot of the pencil-beam-wise range differences between the planning CT and the BD-CT with five BDs for a patient of the HIT data. The difference is color-coded.

## 4.8. Uncertainty Sampling

Simulation of uncertainties during treatment planning is of interest in the field for a long time in the context of plan robustness. A robust plan does not change critically under realistic changes, such as changes in position or range. A good nominal treatment plan is not necessarily a robust plan, therefore simulation of a set of possible uncertainty scenarios ("uncertainty sampling") becomes necessary to evaluate and compare its robustness.

For position changes (i.e. "set-up uncertainties") in photon therapy, this task can be solved under the assumption, that the DD does not change significantly under shift as proposed in Witte et al. (2007), i.e. instead of re-calculating the dose on the CT shifted by a shift vector, the dose can be shifted by the shift vector inversely. This approach is very time efficient, since the plan does not have to be re-calculated, instead only the evaluation metrics have to be calculated for the shifted dose.

For proton therapy, this assumption can certainly not be made and the most straightforward possibility is to re-calculate the nominal DD on each shifted CT (Müller, 2016). In this thesis, only pelvic patient data were analysed, therefore the shift vectors were randomly sampled with a standard deviation of the vector magnitude of  $\sigma = 2 - 3\text{mm}$  following van Herk (2004). Automated tools to sample the shift vectors, do the re-calculation and the analysis was fully implemented in a matRad branch (*dev\_varRBErobOpt*) by Müller (2016), which was used in this thesis for the uncertainty sampling of the proton cases.

For each sampled CT image, the DDs from the set-up uncertainty sampling are analysed using the same analysis metrics as for the nominal CT image. For each analysis metric, the uncertainty samples form a set of values, that is visualized in this thesis using boxplots, given the interquartile range (IQR) as the outer box, and the median value as the central line.

In addition to set-up uncertainties, also radiological depth (rD) uncertainties (cf. section 2.3) have been simulated. For this purpose, the rD for each pencil-beam was up- or down-scaled by 1.5 % consistently over the underlying image while re-calculating the nominal dose plan. For the proton cases, this is also implemented in the uncertainty sampling tool in matRad.

For visualization of the impact of the rD uncertainties on the DD, vertical lines, similar to error bars, are included in the analysis metrics plots.

Uncertainty sampling is performed for each pCT in this thesis. For the BD-CTs and the sCTs, only rD uncertainty sampling is performed, as the set-up uncertainty is the uncertainty which ideally will be strongly reduced by the use of MR image prior or during treatment.

The uncertainty sampling of the pCT images additionally have a special assignment: they are used as an "acceptance band" of analysis metrics for the BD-CTs and sCTs. The sum of the IQR of set-up samples and the range of rD samples create a range of



analysis metric values, which are considered acceptable for the BD-CTs. An example showing the acceptance band in green, the IQR as a boxplot, the rD range as an error bar for arbitrary data in figure 4.13.

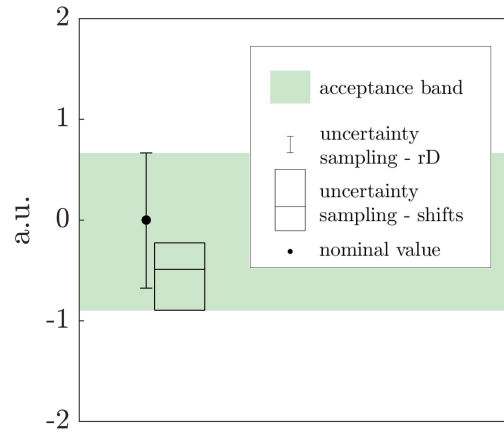


Figure 4.13.: The acceptance band shown in green, created from the range of the boxplot and the errorbar representing the uncertainty sampling for shifts and radiological depth (rD), respectively.



## 5. Results

### 5.1. BD-CTs for photon therapy

The following section summarizes the results of the histogram analysis used to define the bulk density (BD) overwrite values for the bulk density computed tomographies (BD-CTs) derived from the planning computed tomography (pCT) images of the MR guidance dataset. Subsequently, these BD-CTs are analysed regarding their radiological depth (rD) and dose volume histogram (DVH) differences with respect to the pCT, and gamma analysis. In the discussion, the results are evaluated in consideration of the extension to proton therapy, and set into context to literature values.

#### Histogram analysis

The results of the peak and threshold fitting (cf. section 4.5.2) of the relative electron density (rED) histogram are summarized in table 5.1. The mean rED of the fat peak position  $\mu_{fat}$  over the six patients is  $0.934 \pm 0.008$ . The respective value of the soft tissue peak position  $\mu_{st}$  is  $1.027 \pm 0.005$ . The mean threshold rED value is  $0.973 \pm 0.004$ .

Table 5.1.: Mean values  $\mu$  of the two-termed Gaussian function fitted to the relative electron density histogram of the planning CT for each patient of the MR guidance dataset. Each Gaussian peak represents one of the two dominant tissue classes fat and soft tissue (st). The minimum between the two peaks is used as a threshold to separate the two classes.

rED	$\mu_{fat}$	$\mu_{st}$	threshold
MRG1	0.931	1.019	0.970
MRG2	0.948	1.029	0.980
MRG3	0.933	1.029	0.974
MRG4	0.932	1.034	0.973
MRG5	0.924	1.026	0.970
MRG6	0.936	1.026	0.972

The values of  $\mu_{fat}$  and  $\mu_{st}$  from the table are used as overwrite values for the BD-CT images as summarized in section 4.5.1. The BD-CTs are labelled according to the number of BD values that they contain. In total, four BD-CTs are created: BD-CT<sub>WE</sub>, BD-CT<sub>1t</sub>, BD-CT<sub>2t</sub>, and BD-CT<sub>4t</sub>.

## Radiological Depth

The rD analysis is a first indication where the BD-CTs can show strong dosimetric deviations from the pCT. The rD was analysed as explained in equation 2.12 along the pencil-beams used for dose calculation of each beam. The rD is measured from the entrance of the beam into the patient to the point where the beam exits the planning target volume (PTV). Thus, the difference in rD between the pCT and the BD-CT for each beam in the treatment plan can be visualized in two-dimensional (2D) plots. Examples for difference plots are shown in figure 5.1 for BD-CT<sub>WE</sub>, BD-CT<sub>1t</sub>, BD-CT<sub>2t</sub>, and BD-CT<sub>4t</sub> for patient MRG1 for the beam at 0° each. Patient MRG1 was treated for rectum carcinoma with seven beams in total. For the BD-CTs with one or two BD values, anatomical features such as the pelvic bones or an air bubble in the intestine cause large negative or positive differences, respectively. In contrast, for the BD-CT<sub>4t</sub>, the overall differences decreased in a way hardly any of these features are distinguishable.

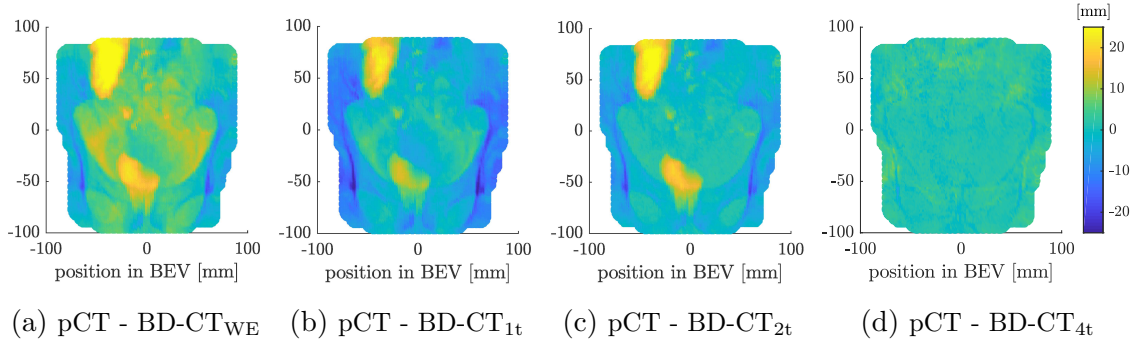


Figure 5.1.: 2D plots of the pencil-beam-wise radiological depth differences between the pCT and (a) BD-CT<sub>WE</sub>, (b) BD-CT<sub>1t</sub>, (c) BD-CT<sub>2t</sub>, and (d) BD-CT<sub>4t</sub>, respectively, for patient MRG1 for the beam at 0°.

The mean absolute differences over all pencil-beams of all beams were compared among the different BD-CTs as summarized in table 5.2. The BD-CT<sub>1t</sub> yields the largest mean absolute differences for all patients and the largest standard deviations for four patients. It performs generally worse than the BD-CT<sub>WE</sub>. The BD-CT<sub>4t</sub> shows the smallest absolute differences with the smallest standard deviations for all patients. Also the variation among the patients is much smaller, giving more reliable results compared to, for example, the BD-CT<sub>2t</sub>, for which mean absolute range differences between 2.8 mm and 7.3 mm occur.

Table 5.2.: Mean and standard deviation (sd) of the absolute difference in radiological depth between the pCT and the BD-CTs (cf. table 4.3) for all pencil-beams of the treatment plan for the six patients of the MR guidance dataset. The average over the six patients is shown in the last row.

patient	mean (sd) absolute rD difference [mm]							
	WE		1t		2t		4t	
MRG1	7.0	(5.2)	11.5	(7.1)	5.9	(4.9)	2.3	(1.9)
MRG2	6.4	(5.9)	7.6	(5.9)	5.5	(6.1)	2.0	(1.6)
MRG3	8.9	(8.6)	10.8	(8.0)	7.3	(8.0)	2.2	(1.9)
MRG4	4.2	(2.7)	7.8	(3.4)	2.8	(2.1)	2.1	(1.5)
MRG5	4.2	(3.2)	5.9	(4.5)	4.6	(3.7)	1.2	(1.1)
MRG6	4.7	(4.7)	7.2	(5.4)	3.9	(4.5)	1.6	(1.6)
average	5.9	(5.1)	8.4	(5.7)	5.0	(4.9)	1.9	(1.6)

## Gamma Analysis

Three-dimensional (3D) gamma analysis (cf. equation 4.5) was performed comparing the dose distribution (DD) from the BD-CTs against the DD from the pCT. In general, "shadows" of failure occur behind intestinal or rectal gas for all the BD-CTs without a BD for air. An example is shown in figure 5.2 for patient MRG2 for the gamma analysis of  $DD_{2t}$  compared to  $DD_{pCT}$ . Similar behaviour is observable more dominantly behind long paths of the beam through bone, e.g. in the pelvic bones. This is also visible in figure 5.2 in the top left corner of the PTV. All DDs from BD-CTs without air and bone BDs show strong influence of this behaviour. For the  $DD_{4t}$ , the air shadow is absent. However, a slight shadow behind cortical bone is still visible due to the fact that only a single BD for the whole bone was assumed, which does not reflect the rD after long paths through cortical bone.

The gamma pass rate (GPR) was analysed for the PTV volume. A summary is shown in figure 5.3, with the mean GPRs of the DD from the BD-CTs of all patients shown as symbols (dots), and the range of GPR values from the six patients as errorbars. GPRs of 100 % were achieved for the  $DD_{4t}$  for all patients. For the  $DD_{2t}$ , the lowest GPR is 96.2 % for patient MRG3. The  $DD_{1t}$  shows the worst mean GPRs, while  $DD_{WE}$  shows a mean GPR of 98 %. The overall lowest GPR was observed for patient MRG4 for the  $DD_{1t}$  (53 %).

## Dose Volume Histogram Analysis

The analysis of the DVH curves includes the analysis of the  $DD_{pCT}$ ,  $DD_{BD-CTs}$ , the uncertainty sampling for range and set-up uncertainties for the pCT, and the

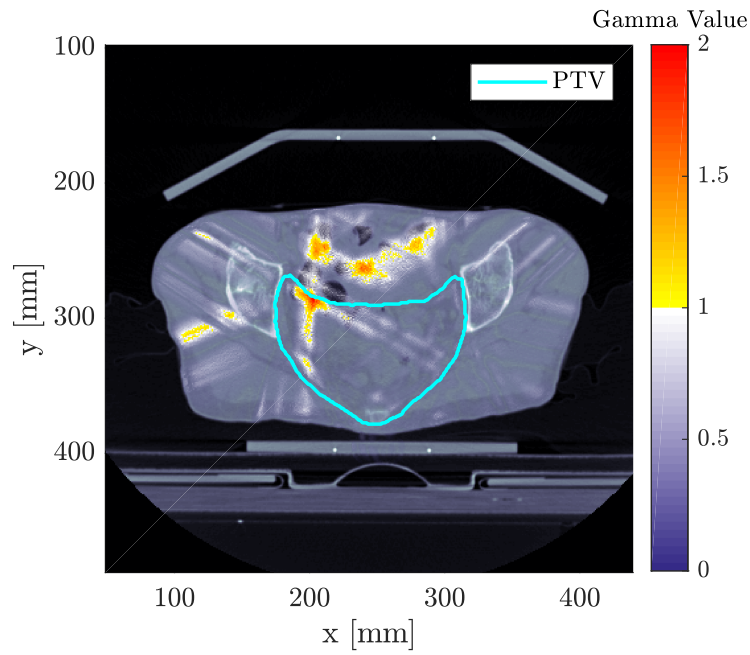


Figure 5.2.: Two-dimensional example of the gamma analysis performed using 2mm/2% criteria for  $DD_{2t}$  of patient MRG2 overlaid onto the corresponding CT slice.

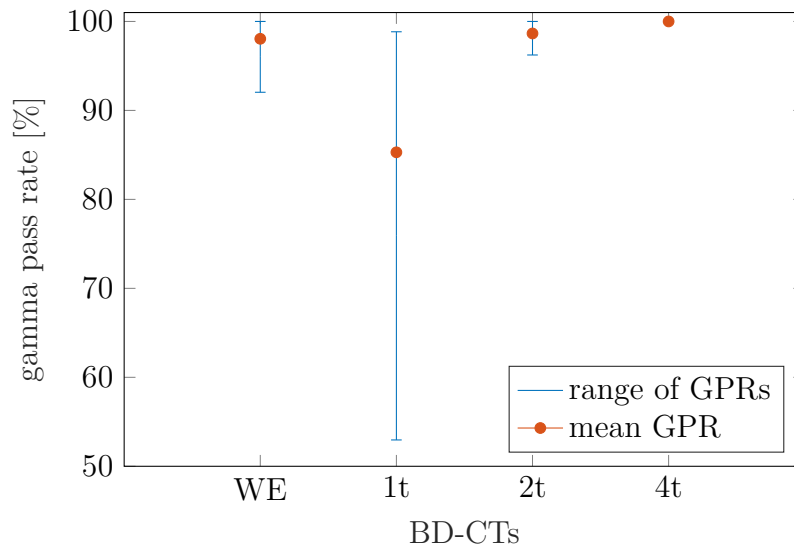


Figure 5.3.: Mean gamma pass rates (GPRs) using a 2mm/2% criterion for the PTV of the  $DD_{BD-CT}$  (dots). The ranges (minimum to maximum GPRs) are shown as errorbars. For the  $DD_{4t}$ , the errorbar is hidden under the dot-marker.

rD uncertainty sampling for the BD-CTs. The set-up uncertainty sampling of the BD-CTs was not included, as this uncertainty would be strongly reduced using the magnetic resonance (MR) prior or during the treatment. The DVHs are analysed for the target volume (the PTV) and the most relevant organs at risk (OARs): the urinary bladder, the spinal cord and the bowel.

For the quantitative analysis of the DVH curves, the dose volume metrics (DVMs)  $D_2$ ,  $D_{50}$  and  $D_{98}$  were evaluated for the PTV, and the  $D_2$  and  $D_{15}$  for the OARs. Table 5.3 shows the mean relative DVM differences for the PTV, the urinary bladder, the spinal cord and the bowel. The standard deviation among the patients is shown in brackets.

Table 5.3.: Mean relative dose difference and standard deviation (sd) of the six patients of the MR guidance dataset for the four  $DD_{BD-CTs}$ .

DVM		mean (sd) relative DVM differences [%]							
		WE		1t		2t		4t	
PTV	$D_2$	0.45	(1.36)	-0.70	(1.16)	-0.13	(1.26)	0.18	(0.29)
	$D_{50}$	-0.35	(0.45)	-1.55	(0.38)	-0.58	(0.25)	0.24	(0.24)
	$D_{98}$	-0.28	(0.30)	-1.43	(0.45)	-0.38	(0.43)	0.30	(0.46)
urinary bladder	$D_2$	-0.97	(1.11)	-2.20	(1.08)	-1.11	(1.02)	0.25	(0.34)
	$D_{15}$	-0.86	(0.93)	-2.06	(0.96)	-2.62	(3.83)	-1.27	(3.36)
spinal cord	$D_2$	-0.15	(0.76)	-1.88	(1.87)	-0.89	(1.87)	1.05	(2.56)
	$D_{15}$	-0.16	(1.34)	-1.90	(2.63)	0.12	(4.35)	2.11	(4.19)
bowel	$D_2$	0.84	(1.21)	-0.29	(1.05)	-1.57	(5.37)	-1.53	(4.21)
	$D_{15}$	0.66	(1.05)	-0.60	(1.55)	-0.07	(1.49)	0.14	(0.27)

None of the BD-CTs yielded DVMs that are unacceptable with respect to the QUANTEC dose constraint recommendations (Marks et al., 2010).

For the PTV, the  $DD_{4t}$  shows the overall smallest differences among the BD-CTs. For  $D_2$ , the standard deviation of the  $DD_{4t}$  is also significantly smaller than for the other BD-CTs. Individual relative DVM difference values might be smaller, for example  $D_{98}$  for the  $DD_{WE}$ , but overall the  $DD_{4t}$  shows best results for the PTV.

Among the OARs, mostly a decrease in dose is observed as indicated by the negative relative differences. Exceptions can be observed for  $DD_{4t}$  for all OARs, for  $DD_{WE}$  for the bowel, and for  $DD_{2t}$  for the spinal cord. The overall largest relative differences are observed for  $DD_{1t}$ , which generally performs worse than  $DD_{WE}$ .

A decrease in PTV dose does not necessarily correspond to an increase in OAR dose. However, it can be expected that an increase in dose may be observed in structures that have not been investigated here.

Figure 5.4 shows the DVMs for the PTV for patient MRG3. This patient was selected for visualization, as it shows the worst results when comparing the DVMs to

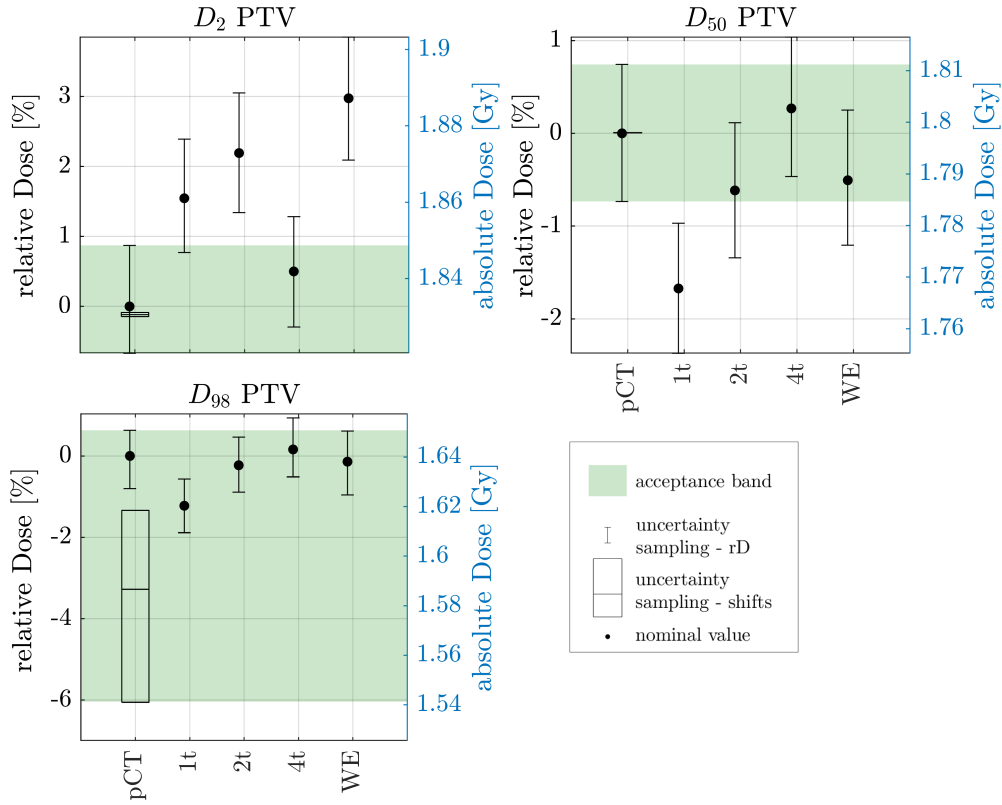


Figure 5.4.: DVMs for the  $DD_{pCT}$  and  $DD_{BD-CTs}$  (circles) for the PTV for one fraction for patient MRG3. The result of the set-up uncertainty sampling for the pCT is shown as a boxplot, while the range of the rD uncertainty sampling for all images is shown as the errorbars.

the acceptance band (cf. section 4.8) among all patients, and the largest relative deviations from the reference DD. The DVMs for the  $DD_{BD-CTs}$  differ up to a maximum of 3 % for the  $D_2$ ,  $DD_{WE}$  has the largest deviations here. For  $D_{50}$  and  $D_{98}$ ,  $DD_{1t}$  shows the largest deviations to the  $DD_{pCT}$ . The observed increase in  $D_2$  and decrease in  $D_{98}$  (for example for  $DD_{1t}$ ) corresponds to a flattening of the DVH curve around the step gradient region. The DVMs for  $DD_{4t}$  lie within the acceptance band in all three cases, whereas for the other BD-CTs, exceedings of the acceptance band can be observed for at least one DVM.

The set-up uncertainty sampling for the pCT shows that a slight error in positioning leads to dose "cold spots" in the PTV, i.e. a partial volume of the PTV does not receive the prescribed dose any more. The result is that the interquartile range (IQR) of the  $D_{98}$  values lies below the nominal value. For  $D_2$  and  $D_{50}$ , the width of the acceptance band is dominated by the rD uncertainty sampling. The influence of the rD uncertainty sampling on the DVMs is consistent between the pCT and the BD-CTs, however shifted by the underlying difference of the DVMs between pCT and BD-CTs. This observation is true for all other patients and structures.



As an example for an OAR, figure 5.5 shows the DVM data for the bowel of the previously showcased patient. The  $DD_{WE}$  shows the largest differences for  $D_2$ . The

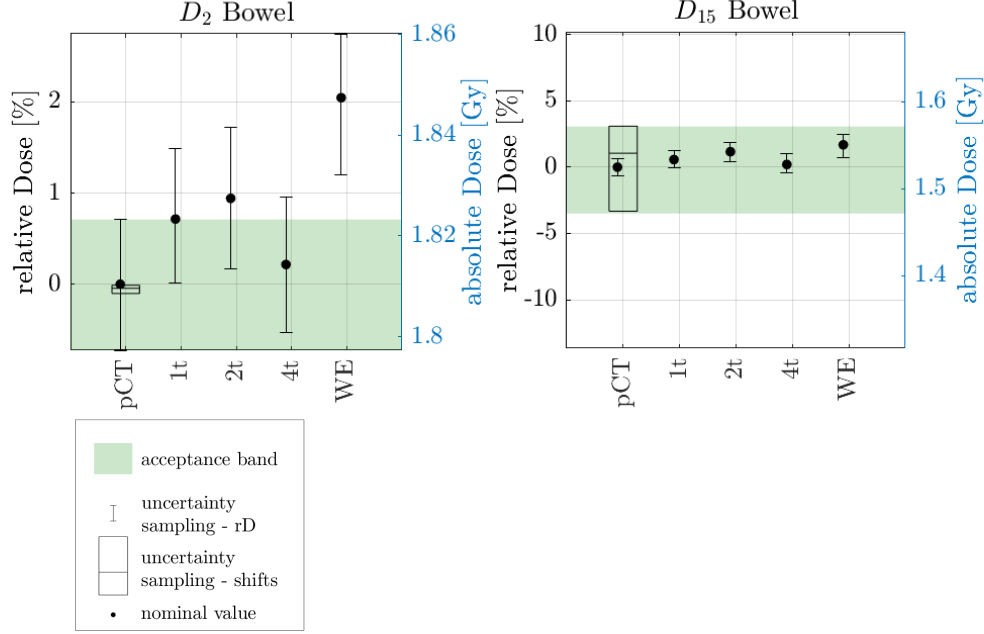


Figure 5.5.: DVMs for the  $DD_{pCT}$  and  $DD_{BD-CTs}$  (circles) for the bowel for patient MRG3. Set-up uncertainty sampling is shown as boxplots, the range of the rD uncertainty sampling is shown as the errorbars. The absolute dose is shown for one fraction.

width of the acceptance band is, similarly to the PTV, dominated by the rD uncertainty sampling, and not by the set-up uncertainty sampling. For  $D_{15}$ , however, the set-up uncertainty sampling dominates. The relative DVM differences for all BD-CTs lie within this acceptance band, including the rD uncertainty sampling of the BD-CTs.

For the PTV, the acceptance band is met for the  $DD_{4t}$  for all patients. For this BD-CT, also the DVMs of the OARs lie within the acceptance band (or below, as lower dose is never critical for OARs) for all investigated cases. The rD uncertainty sampling, visualized as the errorbar, for  $DD_{4t}$  slightly exceeds the acceptance bands, however these "tips" of errorbars which exceed the acceptance bands are worst case scenarios of the rD uncertainty. Most of the errorbar range is within the acceptance band.

For the  $DD_{2t}$ , the acceptance ratio for the PTV drops, as two patients exceed the acceptance band. For three patients, the  $D_2$  value for the bowel lies above the acceptance band.

For  $DD_{1t}$ , all patients fail the PTV acceptance band criterion due to a general underdosage. The acceptance ratio of the OARs is comparable to the one for  $DD_{2t}$ : two failures of  $DD_{1t}$  for the bowel.

DD<sub>WE</sub> fails the PTV criteria for two patients, for which also the OARs show overdosage with respect to the threshold.

The further the DVM value lies at the edge of the acceptance band, the more of the rD uncertainty sampling range, i.e. the errorbar, lies outside of the acceptance band. This is due to the comparable impact of rD uncertainty sampling on the different computed tomography (CT) images, shifted by the underlying difference of the nominal DVM difference.

### 5.1.1. Discussion

**BD values** The standard deviation of the patient-specific mean rED values  $\mu$  and the separation thresholds are below 1 %. This could motivate the usage of cohort-averaged rED values for the BD overwrite values, which in turn would be advantageous in MR-only treatment planning, when no patient-specific information is available. As the number of patients included in this study was low, more patient data would be necessary for more significant cohort results. Compared to White et al. (2016), the average  $\mu_{fat}$  is slightly smaller than the published value, which most probably results from differences in the applied Hounsfield look-up table (HLUT).

**Comparison of the BD-CTs** The analysis of GPRs in this study was performed for the PTV. The gamma analysis is only partially suitable for OARs, since deviations from reference dose are penalized in both directions. Underdosage is not critical for an OAR, but this cannot be distinguished from critical overdosage in the GPR. Thus, gamma analysis was only used for the target volume, where any deviation is of disadvantage for the patient. The results show a clear dependency of the GPR to the granularity of the BD-CT: while the BD-CTs with one BD (WE or 1t) can have GPR down to almost 50 %, the GPRs for the BD-CT<sub>4t</sub> are at 100 % for all patients.

In comparison to the GPR, the 3D cubes showing the gamma index for each voxel are more meaningful to directly correlate the index to anatomy. The observations linking, for example, cortical bone to shadows of gamma index failure are comprehensible based on the strong rD differences in these areas. Comparing the BD-CT<sub>2t</sub> and BD-CT<sub>4t</sub> (and their gamma cubes) clearly shows the improvement in dosimetric agreement due to the bone BD, as the shadows behind bone decrease significantly. The impact of the bone alone might be dominant, but it can not be distinguished quantitatively from the impact of the BD for air, which is also added in the BD-CT<sub>4t</sub>. The absence of air in the BD-CTs is equivalent to the clinical situation, when an intestinal gas bubble moves out of the beam path from the time of the pCT to the actual time of the treatment. The opposite might also occur: a gas bubble moving into the beam during treatment which was not present in the pCT. Both events are currently taken into account by the PTV, to ensure clinical target volume (CTV)

coverage. If the gas bubble is considered an indicator for movement of the bowel, it might however be reasonable to include also a BD for air.

The results of the DVMs show the same tendencies of agreement with the pCT (BD-CT<sub>4t</sub> better than BD-CT<sub>2t</sub>, better than BD-CT<sub>WE</sub>, better than BD-CT<sub>1t</sub>) as the analysis of the mean rD in table 5.3. It might indicate that the standard deviation of the rD difference is not as important as the mean, because both the mean difference of the rD and the mean relative dose differences are comparable between the BD-CT<sub>4t</sub> and the BD-CT<sub>WE</sub>, whereas the BD-CT<sub>WE</sub> has a larger standard deviation.

Even for the best performing BD-CT, BD-CT<sub>4t</sub>, the rD uncertainty sampling lead to a slight exceeding of the acceptance bands at the tips of the errorbars. As for the sampling only a general under- and overestimation of  $\pm 1.5\%$  was assumed, dedicated effort should be invested in the determination of the BD overwrite values to reduce this uncertainty. Possible options are improved imaging techniques, for example dual-energy x-ray CT (DECT) imaging, or direct measurement of rED of tissue. As this uncertainty is even more relevant for proton therapy, more details on possible solutions can be found in the discussion in section 5.2.3.

**Comparison to literature** Eilertsen et al. (2008) presented relative dose differences for the CTV, the bladder and the rectum for synthetic computed tomography (sCT) images with a BD approach for ten prostate cancer patients. They achieved best results for a sCT with one BD for bone equivalent to mass density  $1.3 \text{ g/cm}^3$ , and setting the remaining body to a water-equivalent (WE) BD. Averaged over all patients, they found a relative difference of the mean dose of the CTV of  $0.2\%$ , and a relative difference of the maximum dose of the bladder of  $4.1\%$ . Taking into account that the CTV is less sensitive to dosimetric changes as it is "shielded" by the PTV, even the BD-CT<sub>2t</sub> shows comparably good results for the  $D_{50}$  as a substitute for  $D_{mean}$  in table 5.3. The  $D_2$ , as a substitute for maximum dose, for the urinary bladder performs better than the sCT images in the same study. If you theoretically add the uncertainty of the application of the MR image to the results of this study, the BD-CT<sub>4t</sub> may be an adequate choice of BD overwrites for sCT images. The BD-CT is hardly distinguishable from the pCT any more.

A more comparable study in terms of BD composition of the sCT images was published by Tyagi et al. (2017) for prostate cancer patients. They performed a dosimetric evaluation of sCT images generated with a commercial software. The sCT images contain five BD classes including air, fat, no-fatty soft tissue, soft bone and hard bone. They reported median relative dose differences to the pCT images for 20 patients in the range of  $\pm 0.5\%$  for all investigated structures. Comparing the reported  $D_{95}$  and  $D_{max}$  for the PTV to  $D_{98}$  and  $D_2$  in this study, the results are in good agreement. The published results generally seem slightly better than in this study (where relative dose differences of more than  $0.5\%$  have occurred). For the small and large bowel, they report the IQRs of the relative dose differences of

the maximum dose to be within the -0.5 % to +0.5 % range. However, is it hard to evaluate dosimetric changes if only the maximum dose to a structure is given. The separation of bone into soft and hard bone in Tyagi et al. (2017) might be responsible for the observed smaller dose differences therein.

**Consequences for proton therapy** The metrics used for analysis, i.e. the range analysis, the GPR and gamma cubes, and the analysis of the DVHs, have shown consistent results with complementary information. The predictive power of the rD or range analysis, which could make full dose (re-)calculation redundant, should be subject of future investigations. For now, the analysis metrics are considered suitable for further investigations for proton therapy.

As proton beams are more susceptible to changes in rD, further separation of the bone segments should be considered. One drawback of the presented approach of granularity reduction in the CT image is, that it is not possible to directly quantify the error of individual BD values. For example: the part of the error caused by the single BD for bone is not distinguishable from the part caused by the other BDs. Thus, it may help to identify the contribution of individual BD values investigating BD-CTs with BDs and original content from the pCT.

## 5.2. BD-CTs for proton therapy

Based on the gained knowledge and the developed workflow for the analysis of BD-CTs for intensity-modulated radiotherapy (IMRT), the impact of BD-CTs on the DD for intensity-modulated proton therapy (IMPT) is investigated in this section. The investigation is split into two parts with two different datasets: the Heidelberg Ion-Beam Therapy Center (HIT) dataset, which contains clinical treatment plans used as references, and the Gold Atlas dataset, for which also sCT images will be created in the next section. The analysis of the histogram peak positions is followed again by the analysis of the BD-CTs including range analysis, gamma analysis, and DVH analysis. In the discussion, a set of recommended tissue classes for the sCT images is presented.

### 5.2.1. HIT dataset

#### Histogram Analysis

The peak positions of the two-termed Gaussian function (cf. equation 4.4) fitted to the stopping power ratio (SPR) histogram, the threshold value, and the mean values of the bone segments are summarized in table 5.4. For the analysis of the BD-CTs for proton therapy, the bone segment was further split into soft and hard bone (cf. section 4.5.1). The average value over all patients ( $\pm$  the standard deviation) for the

Table 5.4.: Peak positions  $\mu_{fat}$  and  $\mu_{st}$  of the two-termed Gaussian function fitted to the SPR histogram, representing fat and soft tissue (st). The minimum between the two peaks is used as a threshold to separate those two classes. The mean SPR values of the pre-segmented bone are given for the soft bone, the hard bone, and all bone segments.

SPR	$\mu_{fat}$	$\mu_{st}$	threshold	soft bone	hard bone	all bone
HIT1	0.937	1.048	0.991	1.136	1.541	1.160
HIT2	0.926	1.029	0.987	1.133	1.519	1.157
HIT3	0.925	1.039	0.981	1.151	1.555	1.193
HIT4	0.935	1.032	0.982	1.157	1.520	1.203
HIT5	0.924	1.031	0.985	1.177	1.509	1.207

fat peak is  $0.929 \pm 0.005$ , for the soft tissue peak  $1.036 \pm 0.007$ , for the threshold  $0.985 \pm 0.004$ , for the soft bone segment  $1.151 \pm 0.016$ , to the hard bone segment  $1.529 \pm 0.017$ , and for all bone  $1.184 \pm 0.021$ .

## Range Analysis

The range analysis, as described in section 4.7, was performed comparing the dose from the pCT and the BD-CTs. As the visualization of range difference maps for every patient and beam would be beyond the scope of the thesis, figure 5.6 shows the difference maps for patient HIT3 for the beam at  $0^\circ$  only. This patient shows the largest differences and variations among the cohort. Subfigure (a) clearly shows

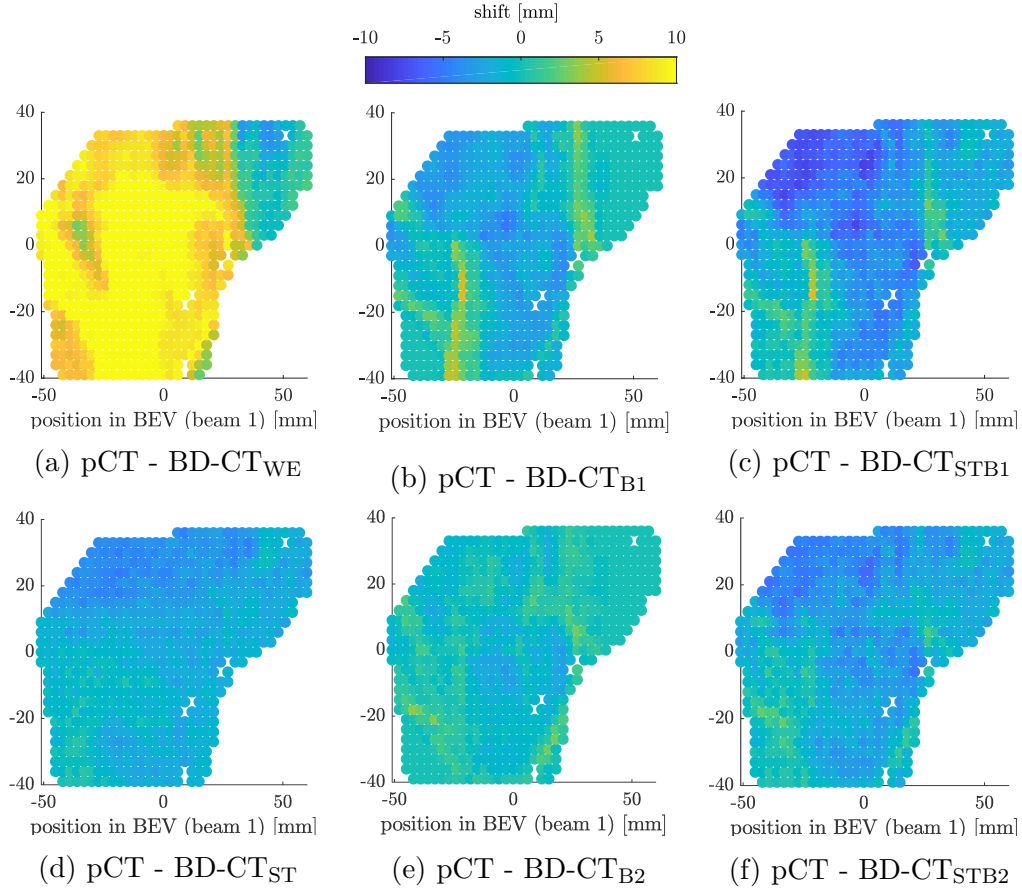


Figure 5.6.: Two-dimensional plots of the pencil-beam-wise range differences between the pCT and (a) BD-CT<sub>WE</sub>, (b) BD-CT<sub>B1</sub>, (c) BD-CT<sub>STB1</sub>, (d) BD-CT<sub>ST</sub>, (e) BD-CT<sub>B2</sub>, and (f) BD-CT<sub>STB2</sub>.

the overshoot of the proton beams using a WE BD-CT. In subfigure (d), the overall difference for BD-CT<sub>ST</sub> is much smaller, though a slowly varying difference can be seen. BD-CT<sub>B1</sub> in subfigure (b) creates larger range differences as compared to BD-CT<sub>B2</sub> in subfigure (e). Those differences are enhanced for BD-CT<sub>STB1</sub> in subfigure (c) and BD-CT<sub>STB2</sub> in subfigure (f), which behave like sums of BD-CT<sub>ST</sub> and BD-CT<sub>B1/B2</sub>.

The pencil-beam-wise absolute range differences are averaged for each patient, the results are summarized in table 5.5. Largest absolute differences and standard de-

Table 5.5.: Mean and standard deviation (sd) of the absolute range difference between the pCT and the BD-CTs (cf. table 4.3) for all pencil-beams of the treatment plan for the five patients of the HIT dataset. The average over all patients is shown in the last row.

patient	mean (sd) absolute range difference [mm]					
	WE	ST	B1	B2	STB1	STB2
HIT1	11.5 (6.1)	2.7 (1.3)	1.7 (1.6)	1.5 (1.1)	3.0 (1.8)	2.4 (1.6)
HIT2	3.5 (2.5)	0.6 (0.8)	1.0 (1.4)	0.7 (1.2)	1.5 (1.5)	1.4 (1.4)
HIT3	8.2 (4.1)	1.9 (1.2)	1.7 (1.4)	0.8 (0.8)	3.2 (2.1)	2.3 (1.5)
HIT4	12.3 (4.5)	1.2 (0.9)	2.0 (1.6)	1.1 (0.9)	2.8 (1.8)	1.6 (1.1)
HIT5	7.3 (3.7)	0.7 (0.6)	1.5 (1.2)	0.9 (0.7)	1.9 (1.4)	1.4 (1.1)
average	8.6 (4.2)	1.5 (0.9)	1.6 (1.4)	1.0 (1.0)	2.5 (1.7)	1.8 (1.4)

variations among the BD-CTs are observed for the BD-CT<sub>WE</sub> for all patients. Going from one BD for bone to two BD, an improvement, i.e. a decrease in difference, can be observed for all patients. Also the standard deviations decrease, which indicates more reliable results. The absolute difference in rD for BD-CT<sub>ST</sub> ranges from 0.6 mm to 2.7 mm. The larger the difference for this BD-CT, the larger also the absolute difference for BD-CT<sub>STB1</sub> and BD-CT<sub>STB2</sub>, respectively. This is due to the combinatory effect of the BDs. Large variations among the patients can be observed, most prominently for the BD-CT<sub>WE</sub>, but with similar trends also for the other BD-CTs.

## Gamma Analysis

Gamma analysis was performed comparing the  $DD_{pCT}$  to  $DD_{BD-CTs}$ . Examples of two 2D slices of the gamma cube comparing  $DD_{pCT}$  and  $DD_{ST}$  for one patient are shown in figure 5.7. In both examples it becomes clear from the figure, that the failures of the gamma criterion mostly occur just outside of the PTV, which corresponds to the end of the proton ranges of the beams. This is the case for all BD-CTs, however local "hotspots" of GPR failure can be observed at various positions. For the  $DD_{B1}$ , hotspots can be observed after significant passages through hard bone (results not shown). The GPR failure in figure 5.7 is present in all BD-CTs with soft tissue BDs.

The GPRs were evaluated for the PTV only. The results of the gamma analysis for all patients are summarized in figure 5.8. The  $DD_{WE}$  shows the lowest GPRs for all patients, visible from the errorbar. The GPRs for the  $DD_{B2}$  are closer to the optimal 100 % than for  $DD_{B1}$  for all patients. For the  $DD_{STB1}$  and  $DD_{STB2}$ , the GPRs show larger deviations due to the combinatory effect of the BDs. The worst GPR for the  $DD_{STB2}$  of 97.6 % was observed for patient HIT1.

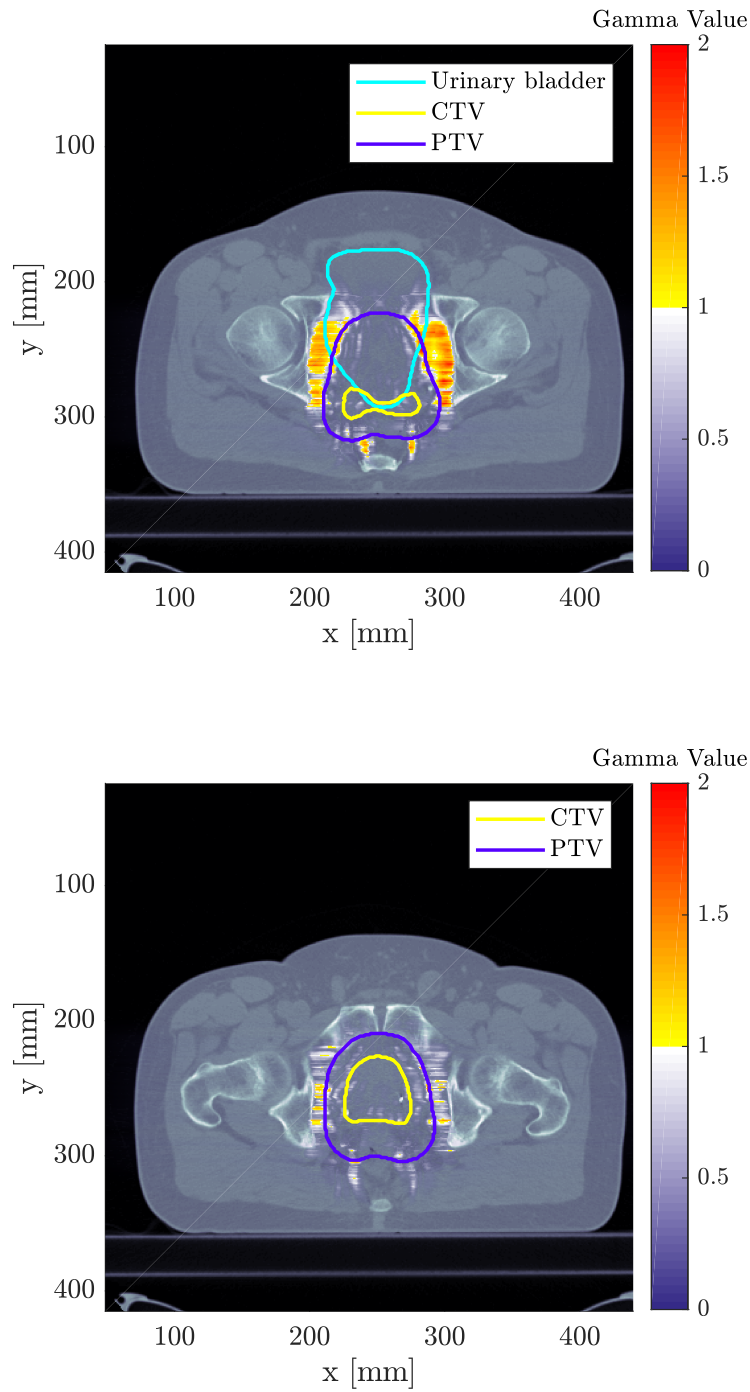


Figure 5.7.: Two-dimensional examples of the gamma analysis performed using a 2mm/2% criteria for  $DD_{ST}$  of patient HIT3 overlaid onto the corresponding CT slices. On top, a slice through the urinary bladder at the edge of the CTV is shown, which is the source of GPR failures left and right to the bladder. At the bottom, a slice through the target volume is shown.



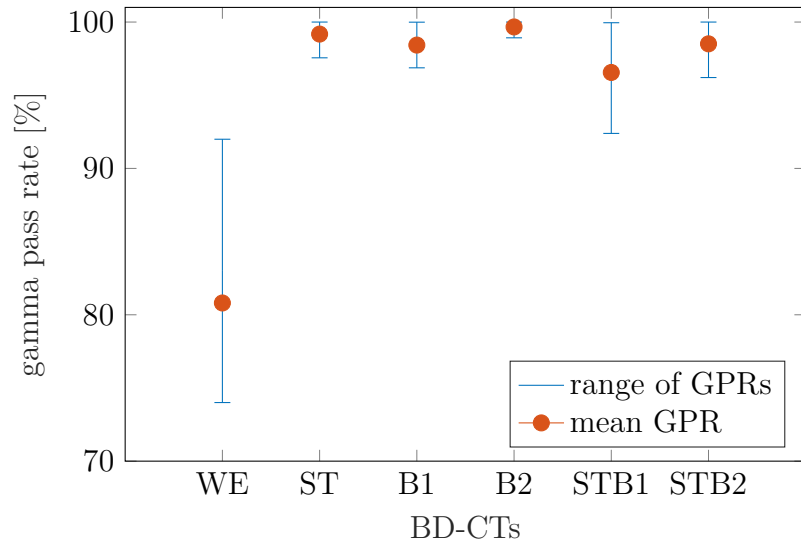


Figure 5.8.: Mean gamma pass rates (GPRs) using a 2mm/2% criterion for the PTV of  $DD_{BD-CT}$  (dots) for the HIT dataset. The ranges (minimum to maximum GPRs) are shown as errorbars.

Table 5.6.: Mean relative dose difference and standard deviation (sd) of the five patients of the HIT dataset for the different  $DD_{BD-CTs}$ .

		mean (sd) relative DVM differences [%]											
DVM		WE		ST		B1		B2		STB1		STB2	
PTV	$D_2$	-0.08	(1.13)	0.03	(0.08)	0.04	(0.09)	0.03	(0.10)	-0.02	(0.06)	-0.01	(0.06)
	$D_{50}$	-0.88	(0.36)	-0.02	(0.04)	0.00	(0.03)	-0.04	(0.10)	0.03	(0.15)	0.01	(0.06)
	$D_{98}$	2.43	(13.68)	-1.57	(2.59)	-0.88	(1.83)	-2.27	(4.13)	-1.10	(1.72)	-1.56	(1.27)
urinary bladder	$D_{max}$	0.30	(2.00)	0.42	(0.37)	-0.03	(0.05)	-0.02	(0.09)	0.06	(0.27)	0.07	(0.24)
	$D_{15}$	3.86	(10.24)	-4.97	(4.69)	0.49	(1.52)	-0.25	(3.10)	-3.47	(3.60)	-2.33	(2.54)
rectum	$D_{max}$	1.01	(1.22)	1.07	(1.50)	0.53	(0.62)	0.56	(1.24)	1.07	(1.22)	0.81	(1.00)
	$D_{15}$	1.34	(4.50)	0.66	(0.54)	0.05	(0.27)	0.29	(0.34)	0.57	(0.48)	0.50	(0.47)
RPW	$D_{max}$	34.57	(55.59)	-2.28	(3.49)	1.26	(3.28)	1.29	(3.26)	-2.64	(3.25)	-2.46	(3.23)
	$D_{15}$	9.37	(6.72)	-0.39	(1.26)	0.55	(1.24)	-0.09	(0.36)	0.31	(2.19)	-0.18	(1.71)

## Dose Volume Histogram Analysis

For the HIT dataset, the PTV and the most important OARs used for dose constraints during plan optimization, including the urinary bladder, the rectum and the rectal posterior wall (RPW), have been analysed. The mean relative differences of the DVMs of the five patients are summarized in table 5.6 with the standard deviation in brackets. In proton treatment planning for prostate cancer,  $D_{max}$  is used as a dose constraint for the urinary bladder and the rectum. In both cases, it should not exceed the prescribed dose (i.e.  $D_{max} < 100$  %). For this reason,  $D_2$  was replaced with  $D_{max}$  from here on for the analysis of the OARs. For the RPW,  $V_{50\text{Gy}(RBE)} < 17$  % and the  $V_{32\text{Gy}(RBE)} < 35$  %, i.e. the partial volume of the structure that receives 50 Gy(RBE), and 32 Gy(RBE), respectively, are used as constraints during plan optimization. However, in all cases in this study, the  $V_{50\text{Gy}(RBE)}$  and the  $V_{32\text{Gy}(RBE)}$  are well below 1 % and thus are not reported explicitly. For the PTV, the largest deviations from the reference DD were observed for the  $D_{98}$ . A negative  $D_{98}$  difference, as observed for  $DD_{ST}$ ,  $DD_{B1}$ ,  $DD_{B2}$ ,  $DD_{STB1}$ , and  $DD_{STB2}$ , indicates a worse target coverage. Surprisingly,  $DD_{B2}$  shows even worse target coverage than  $DD_{B1}$ . In addition, the variation among the patients is also larger, indicated by larger standard deviations. Similar trends are also present among the OAR DVMs.

$DD_{WE}$  shows the largest dose differences for the RPW of up to 34.57 % for  $D_{max}$ . At the same time,  $D_{max}$  of the pCT is much smaller (between 0.5 Gy(RBE) and 1 Gy(RBE)) for the RPW than for the bladder or the rectum, leading to larger relative differences. For the other structures,  $DD_{WE}$  does not show significantly larger differences in comparison to the other DDs.

$D_{max}$  increases for the urinary bladder and the rectum, which might hint towards a exceeding of the dose constraints. For the RPW,  $D_{max}$  decreases with respect to the reference DD for most patients, dominated by the strong impact of the soft tissue BDs.

Figure 5.9 shows the DVMs for the PTV for patient HIT5 as an example. The relative dose differences of all BD-CTs are below 0.5 % for the shown DVMs. Even for  $BD-CT_{WE}$ , which shows the largest deviations, the deviations are smaller than 0.5 %. The deviations can thus be considered negligible.

The acceptance bands are influenced both by the rD and the set-up uncertainty sampling. Only for  $D_{98}$ , a clear dominance of the set-up uncertainty sampling is observable.  $BD-CT_{WE}$  exceeds the acceptance bands for  $D_2$  and  $D_{50}$ . For  $D_{98}$ , also exceedings of the other BD-CTs can be observed in the order of 0.1 percent points. However, as the overall deviations are so small, these deviations are still negligible. The errorbars of the BD-CTs, representing the rD uncertainty sampling for these images, also lie within the acceptance bands for most parts. The tips of the errorbars, however, might exceed the acceptance bands slightly, even though the nominal DVM values lies well within the acceptance band.

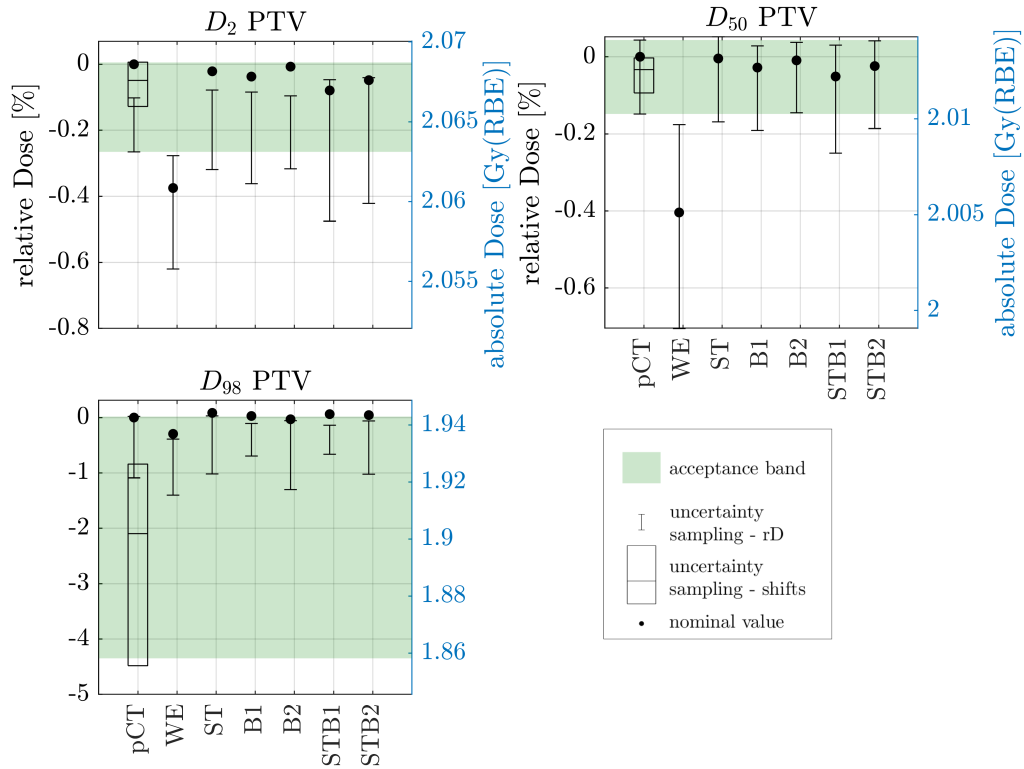


Figure 5.9.: DVMs for the pCT and BD-CTs for the PTV for patient HIT5.

For the remaining patients not shown here, the results for the analysis of  $D_2$  and  $D_5$  are similar and the relative dose differences are small (below 0.5 %, except for BD-CT<sub>WE</sub>). Only for  $D_{98}$ , larger deviations occur.

Similarly to the showcased patient, several other patients show exceedings of the acceptance bands, however the deviations are still small enough in all cases to be considered negligible.  $D_{98}$  is less affected by acceptance band exceedings, as the set-up uncertainty sampling can lead to deviations up to 30 percent points. BD-CT<sub>WE</sub> shows exceedings of the PTV acceptance bands for patients HIT1, HIT2, and HIT3. While the impact of the rD uncertainty sampling between the pCT and the BD-CTs is comparable in width to the shown example, for other patients more variation can occur. This also impacts the overlap between the acceptance bands and the errorbars.

A corresponding plot is shown for the urinary bladder in figure 5.10 also for patient HIT5. The relative deviations of the BD-CTs are well below 1 % for  $D_{15}$  and even below 0.5 % for  $D_{max}$ . An increase in dose is observed for DD<sub>ST</sub>, DD<sub>STB1</sub>, and DD<sub>STB2</sub> to a maximum of approximately 0.4 %. All DVMs, however, lie within the acceptance bands.

For  $D_{max}$ , the rD uncertainty sampling dominates the width of the acceptance band, for  $D_{15}$  it is the set-up uncertainty sampling. Only minor exceedings of the BD-CT errorbars can be observed for  $D_{max}$ .

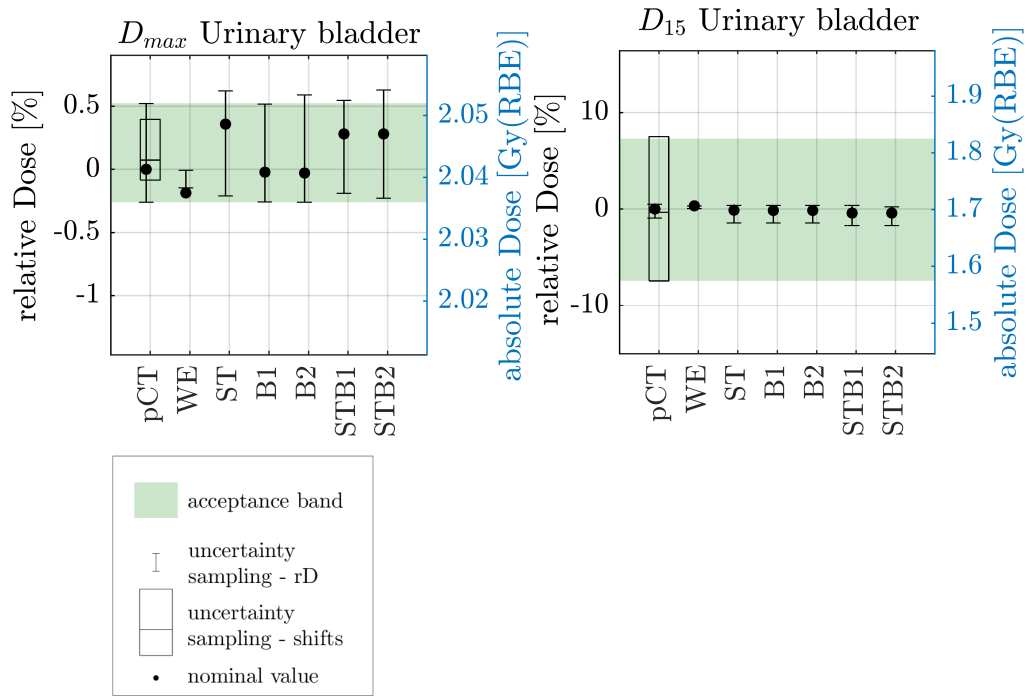


Figure 5.10.: DVMs for  $DD_{pCT}$  and  $DD_{BD-CTs}$  (circles) for the urinary bladder for patient HIT5. Set-up uncertainty sampling is shown as boxplots, the rD uncertainty sampling is shown as errorbars.

For the other patients, most exceedings of the acceptance bands were observed for  $BD-CT_{WE}$  for at least one OAR for patient HIT1, HIT2, and HIT3. A critical exceeding of the acceptance band was observed for patient HIT1 for  $DD_{ST}$  for  $D_{max}$  of the urinary bladder, and for  $DD_{ST}$  and  $DD_{STB1}$  for  $D_{max}$  of the rectum. For all other OARs and DVMs, deviations are well below 1 %. Larger deviations above 1 % occur, but are within the acceptance band.

### Exemplary insertion of a BD for the urinary bladder

In the gamma analysis for the  $DD_{ST}$  shown in figure 5.7, it is clearly visible, that the urinary bladder has a strong impact on the DD. It is difficult to quantify, how much of the dose differences seen for  $DD_{ST}$  is due to the BD for fat and the non-fatty soft tissue, and how much is due to the wrong overwrite of the BD in the bladder. Thus, for one patient (HIT3), the urinary bladder was manually overwritten with a BD value equivalent to the mean SPR value inside the delineated bladder ( $SPR = 0.9995$ ). The grey values of the bladder lie on the left slope of the soft tissue peak (compare to figure 4.5 or 4.6) and purely grey value based segmentation in the image is not possible. Thus, the delineated urinary bladder included in the patient dataset

was used to overwrite the bladder volume with a BD value. The BD-CT created in this way is referred to as  $\text{BD-CT}_{\text{ST+bladder}}$ , the re-calculated dose as  $\text{DD}_{\text{ST+bladder}}$ , respectively.

The range difference improved from 1.9 mm to 1.5 mm for the mean absolute difference, and from 1.2 mm to 0.8 mm for the standard deviation of the absolute difference.

The GPR for the PTV increased from 98.4 % to 99.2 %. Figure 5.11 shows the respective slice of the gamma analysis of  $\text{DD}_{\text{ST+bladder}}$  as in figure 5.7. The gamma

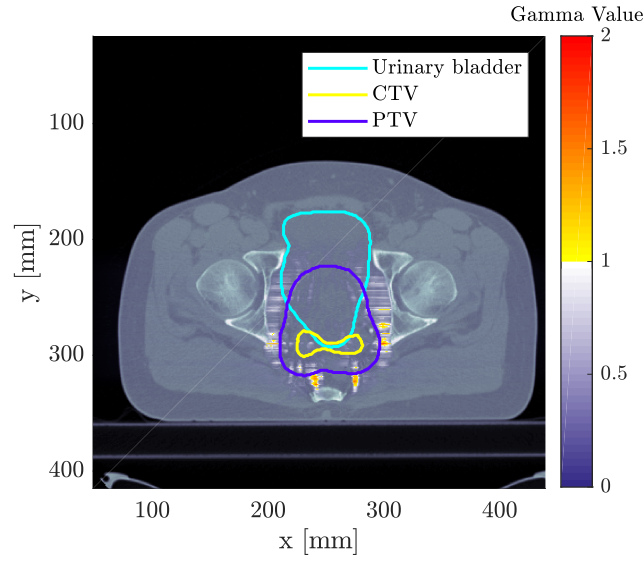


Figure 5.11.: Two-dimensional example of the gamma analysis performed using a 2mm/2% criteria for  $\text{DD}_{\text{ST+bladder}}$  of patient HIT3 overlaid onto the corresponding CT slice. The same slice is shown as in figure 5.7.

value hotspots that were created at the distal ends of beams passing through the bladder are not present any more. The improvement in the GPR is also reflected in the improvement of the relative differences of the DVMs comparing  $\text{DD}_{\text{ST}}$  and the new  $\text{DD}_{\text{ST+bladder}}$  in table 5.7. For all DVMs, the relative difference decreases. The only exception is  $D_{15}$  for the RPW, which slightly increases from 0.13 % to 0.35 %, and can thus still be considered acceptable. Especially for the urinary bladder itself, strong improvement is observed:  $D_{15}$  improves from -7 % to 1 %.

Table 5.7.: Relative DVM differences for  $DD_{ST}$  and  $DD_{ST+bladder}$  for patient HIT3.

	DVM	relative DVM difference [%]	
		ST	ST+bladder
PTV	$D_2$	0.17	-0.09
	$D_{50}$	-0.05	-0.03
	$D_{98}$	-5.66	-2.21
urinary bladder	$D_{max}$	0.56	-0.12
	$D_{15}$	-7.14	0.71
rectum	$D_{max}$	3.66	0.66
	$D_{15}$	0.76	-0.11
RPW	$D_{max}$	-6.59	-4.04
	$D_{15}$	0.13	0.35

### 5.2.2. Gold Atlas dataset

As the HIT dataset only contained 5 patients, among which one was a cervix cancer patient and the others were prostate cancer patients, the results are not particularly significant. In contrast, the Gold Atlas dataset contains 19 patients, for which prostate cancer treatment plans have been optimized.

#### Histogram Analysis

All pCT images from the Gold Atlas dataset have been translated to SPR images using a clinical HLUT, since no clinical HLUTs calibrated for the respective CT scanners have been delivered with the image data. However, no bias is expected from using this HLUT, since the pCT as the gold standard is also translated to SPR using the same HLUT for dose calculation. To account for the different scanners, the BD SPR values derived from the Gaussian fit and the averaging over the bone structures are averaged over the patients of each site. The results are shown in table 5.8.

Table 5.8.: Mean SPR values for the patients of the Gold Atlas dataset. Mean values are shown for each site (GA1-3) and for all patients.

SPR	$\mu_{fat}$	$\mu_{ST}$	threshold	soft bone	hard bone	all bone
GA1.1-8	0.956	1.031	0.996	1.136	1.478	1.159
GA2.1-7	0.943	1.022	0.991	1.138	1.514	1.174
GA3.1-4	0.957	1.041	0.997	1.139	1.478	1.163

Differences between the different sites are less pronounced between GA1 and GA3, whereas GA2 shows larger differences compared to the former two. The mean value

of all patients ( $\pm$  the standard deviation) for the fat peak is  $0.951 \pm 0.008$ . For the soft tissue peak it is  $1.030 \pm 0.009$ , for the threshold it is  $0.994 \pm 0.003$ , for the soft bone segment it is  $1.138 \pm 0.010$ , for the hard bone segment it is  $1.491 \pm 0.026$ , and for all bone it is  $1.165 \pm 0.016$ .

### Range Analysis

The analysis of the pencil-beam-wise absolute range difference between the pCT and the BD-CTs averaged over all 19 patients is shown in table 5.9. Comparing the

Table 5.9.: Mean and standard deviation (sd) of the absolute difference in range between BD-CTs and pCT for all pencil-beams of the treatment plan for the 19 patients of the Gold Atlas dataset, averaged per site and over all patients (bottom row).

site	mean (sd) absolute range difference [mm]					
	WE	ST	B1	B2	STB1	STB2
GA1.1-8	10.7 (2.8)	0.9 (0.8)	1.7 (1.9)	1.1 (1.3)	2.0 (2.3)	1.4 (1.5)
GA2.1-7	9.5 (3.8)	0.8 (0.9)	2.2 (2.4)	1.1 (1.3)	2.1 (2.6)	1.3 (1.7)
GA3-1.4	10.7 (2.7)	0.6 (0.7)	1.9 (1.9)	1.1 (1.2)	2.1 (2.2)	1.3 (1.4)
average	10.3 (3.1)	0.80 (0.8)	1.9 (2.1)	1.1 (1.3)	2.1 (2.4)	1.3 (1.6)

results for the  $\text{BD-CT}_{\text{ST}}$ ,  $\text{BD-CT}_{\text{B1}}$  and  $\text{BD-CT}_{\text{B2}}$ , the impact of the soft tissue BD on the absolute range difference is always smaller than the impact of the bone BD values. This is in contrast to the observations with the HIT dataset, where the soft tissue BDs had a comparable or even stronger impact on the difference. The overall mean absolute differences for the  $\text{BD-CT}_{\text{STB1}}$  and  $\text{BD-CT}_{\text{STB2}}$  are smaller than for the HIT dataset, where averaged over all patients the mean differences were 2.5 mm and 1.8 mm for the two BD-CTs, respectively.

Figure 5.12 shows exemplary range difference plots for patient GA3.3, which shows the largest mean absolute range differences for the  $\text{BD-CT}_{\text{B1}}$  and  $\text{BD-CT}_{\text{B2}}$  among all patients of the dataset, and consequently also very large mean absolute differences for the  $\text{BD-CT}_{\text{STB1}}$  and  $\text{BD-CT}_{\text{STB2}}$ .

In subfigure (a), the same overshooting of the proton beams can be observed as for the  $\text{BD-CT}_{\text{WE}}$  for the HIT dataset. As the mean absolute range difference averaged over the patient cohort already suggested, the difference plot shown in subfigure (d) for  $\text{BD-CT}_{\text{ST}}$  shows only minor variations. The central negative range difference in subfigure (b) for  $\text{BD-CT}_{\text{B1}}$  (corresponding to an undershooting of the proton beam) and the overshoot in the lower left corner smooth out for  $\text{BD-CT}_{\text{B2}}$  in subfigure (e). Subfigures (c) and (f) are visually almost indistinguishable from (b) and (c), respectively, due to the negligible impact of the soft tissue BD compared to the bone BDs.



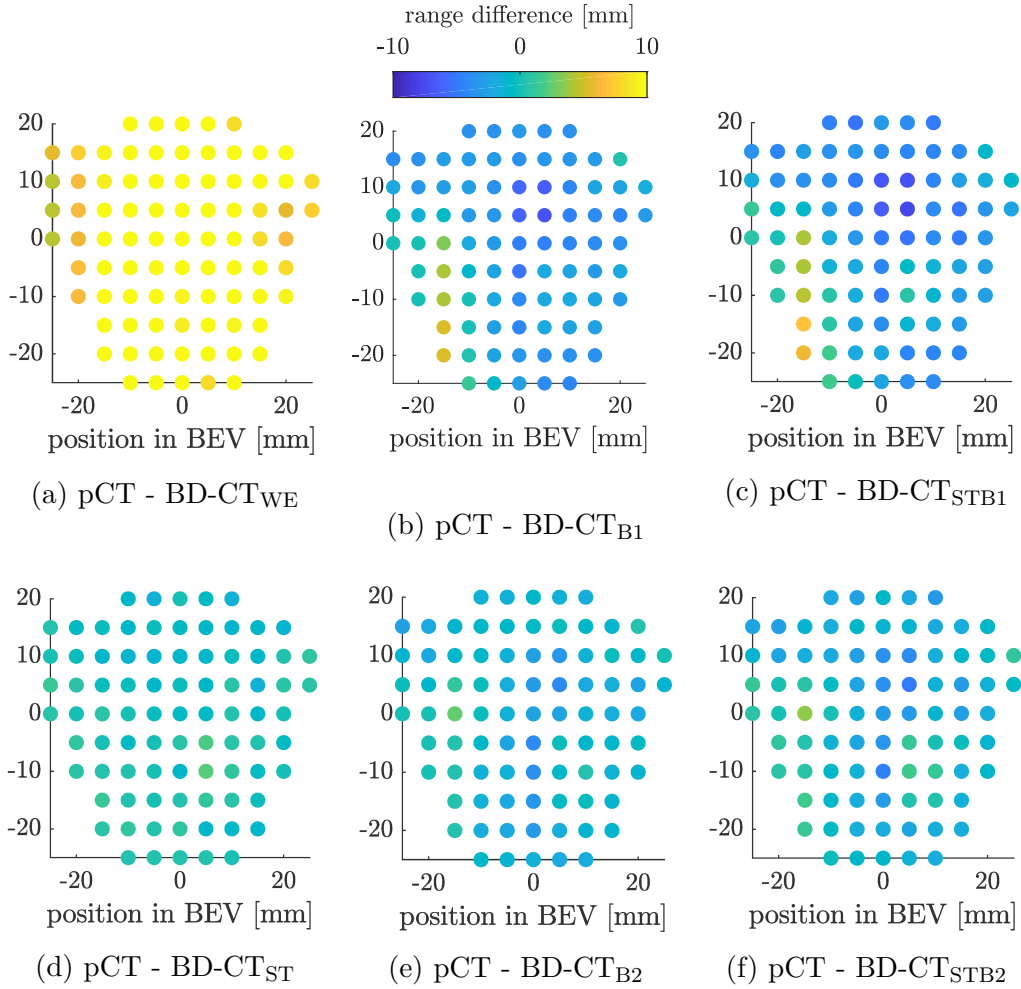


Figure 5.12.: Two-dimensional plots of the pencil-beam-wise range differences between the pCT and the BD-CTs for patient GA3.3 and the beam at  $0^\circ$ .

### Gamma Analysis

Figure 5.13 shows the mean GPRs for the PTV for all 19 patients of the Gold Atlas dataset. The errorbar again represents the range of GPRs observed among those patients. Similar to the observations with the HIT dataset, the DD<sub>WE</sub> shows the lowest GPRs among the BD-CTs. For the DD<sub>ST</sub>, no gamma index failure can be observed, leading to GPRs of 100 % for all patients. The improvement in GPR going from DD<sub>B1</sub> to DD<sub>B2</sub>, and from DD<sub>STB1</sub> to DD<sub>STB2</sub>, is represented by the increase mean GPR and the decrease in range of GPRs. The smallest GPR for the DD<sub>STB2</sub> is 97.1 % for patient GA1.3.

The analysis of the 3D gamma cubes indicates two possible reasons for failures of the DD<sub>ST</sub>, which are, however, outside of the PTV. First, metallic markers in the

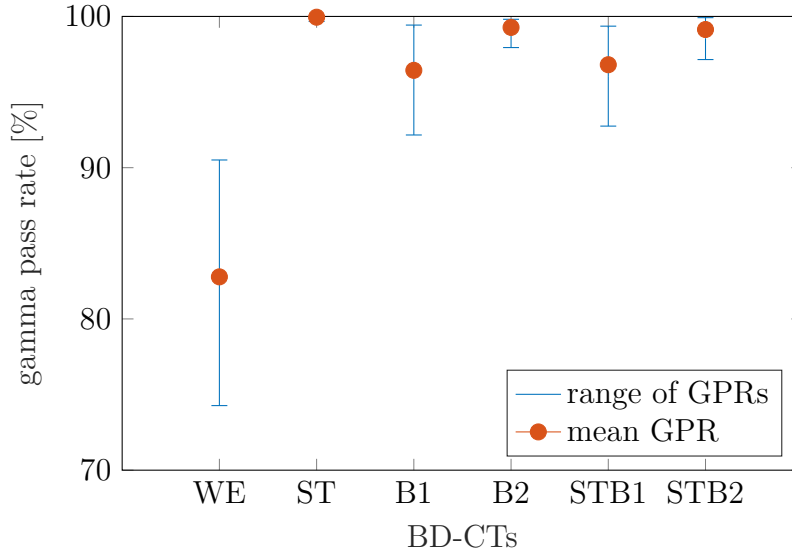


Figure 5.13.: Mean gamma pass rates (GPRs) using a 2mm/2% criterion for the PTV of  $DD_{BD-CT}$  (dots) for the HIT dataset. The ranges (minimum to maximum GPRs) are shown as errorbars.

CT image occurred for two patients of GA1 and two patients of site GA3. In this case, the SPR is overwritten with soft tissue BD values. Second, contrast agent in the urine that settles at the posterior side of the bladder and thus close to the target volume was also observed. The contrast agent increases the SPR of the urine, but is overwritten with a soft tissue BD as well. This occurred for five out of seven patients for site GA2. An example of a 2D slice of the pCT and the corresponding failure of the gamma values for the  $DD_{ST}$  is shown in figure 5.14.

For the  $DD_{B1}$ , at all passages of the beam through hard bone, failures occur in the 3D gamma cube (results not shown). Those are considerably reduced for the  $DD_{B2}$ . However, also within the  $BD-CT_{B2}$ , cortical bone might not be well represented at all times, which leads to smaller "hotspots" of gamma values in the 3D cubes.

## Dose Volume Histogram Analysis

The differences of the DVMs for the  $DD_{BD-CTs}$  in the PTV, the urinary bladder, the RPW and the rectum relative to the  $DD_{pCT}$  are summarized for all patients in table 5.10. Similar to all metrics and datasets analysed before,  $DD_{WE}$  yields the largest, unacceptable deviations for all structures and it is therefore not further discussed.  $DD_{ST}$  shows almost no deviations for the analysed DVMs and structures. The largest deviation is observed for  $D_{15}$  for the urinary bladder with  $(1.03 \pm 4.15) \%$ , which, however, does not exceed clinical constraints in any investigated case. Due to the almost negligible impact of the soft tissue BDs, the DDs of the

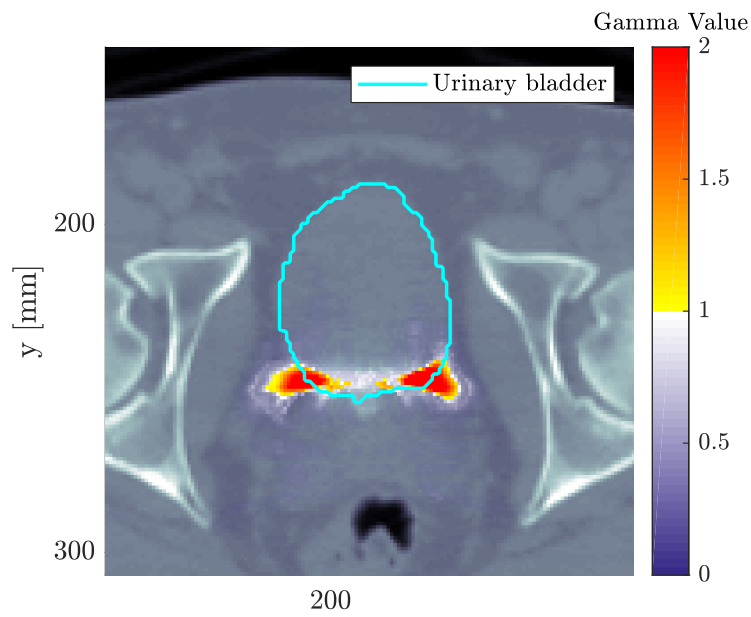


Figure 5.14.: Two-dimensional example of the gamma analysis performed using 2mm/2% criteria for  $DD_{ST}$  of patient GA2.5 overlaid onto the corresponding CT slice. For better visualization of the contrast agent settlement in the urinary bladder, this image section is zoomed in.

Table 5.10.: Mean relative dose difference and standard deviation (sd) of the 19 patients of the Gold Atlas dataset for the different  $DD_{BD-CTs}$ .

		mean (sd) relative DVM differences [%]											
	DVM	WE		ST		B1		B2		STB1		STB2	
PTV	$D_2$	-0.52	(0.31)	0.02	(0.04)	-0.05	(0.08)	0.00	(0.04)	-0.02	(0.10)	0.02	(0.06)
	$D_{50}$	-0.66	(0.34)	0.06	(0.07)	-0.17	(0.07)	-0.05	(0.08)	-0.10	(0.10)	-0.01	(0.09)
	$D_{98}$	0.46	(1.29)	0.19	(0.57)	-4.63	(2.91)	-1.20	(1.55)	-4.14	(3.23)	-1.27	(1.62)
urinary bladder	$D_{max}$	1.00	(1.37)	0.09	(0.22)	0.20	(0.71)	0.28	(0.63)	0.22	(0.75)	0.25	(0.59)
	$D_{15}$	9.97	(5.61)	1.03	(4.15)	-2.88	(1.17)	-0.54	(1.07)	-1.55	(4.30)	0.54	(4.32)
rectum	$D_{max}$	2.24	(1.05)	0.05	(0.13)	0.28	(0.51)	0.13	(0.40)	0.28	(0.54)	0.18	(0.40)
	$D_{15}$	-3.17	(4.02)	-0.28	(0.57)	-0.33	(1.07)	-0.30	(0.85)	-0.78	(1.21)	-0.66	(1.04)
RPW	$D_{max}$	4.98	(4.98)	0.05	(0.59)	0.18	(1.13)	0.30	(1.14)	0.08	(1.06)	0.23	(1.27)
	$D_{15}$	-2.92	(3.33)	-0.19	(0.58)	-0.75	(2.55)	-0.24	(1.45)	-0.88	(2.61)	-0.61	(1.32)

combined BDs for soft tissue and bone ( $DD_{STB1}$  and  $DD_{STB2}$ ) are dominated by the differences observed for the  $DD_{B1}$  and  $DD_{B2}$ .

The PTV is not covered adequately with dose for  $DD_{B1}$  and  $DD_{STB1}$ , as can be seen by the mean decrease of  $D_{98}$  of more than 4 %. This DVM and the  $D_{15}$  for the urinary bladder show the most significant improvements when going from  $DD_{B1/STB1}$  to  $DD_{B2/STB2}$ . For most OARs,  $D_{max}$  slightly increases (maximum of 0.28 %) and  $D_{15}$  decreases (maximum of 2.88 %). Generally speaking, the differences are quite small, with only few DVMs exceeding a 1 %-difference.

Patient GA2.6 was the only patient who showed a failure of the acceptance band criterion for an OAR. Thus, for this patient, the DVMs for the PTV and the urinary bladder are shown in figures 5.15 and 5.16, respectively. For the PTV, the observed relative differences to  $DD_{pCT}$  for the  $D_2$  and  $D_{50}$  are small (below 0.5 %), in agreement with the mean values in table 5.10. For  $D_{98}$ , the relative differences are larger, up to approximately 5 % for  $DD_{B1}$ , however all differences are covered by the acceptance band. As for the HIT dataset, the width of the acceptance band for  $D_{98}$  is dominated by the set-up uncertainty sampling. The errorbars, representing the rD uncertainty sampling, show different behaviour for each BD-CT compared to the pCT. For the PTV, the largest parts of the rD uncertainty sampling are also covered by the acceptance bands.

For the remaining 18 patients Gold Atlas dataset, the situation is similar as for the shown example: for  $D_2$  and  $D_{50}$ , all DVM differences are below 0.2 %, and for  $D_{98}$  larger deviations occur, but are covered by the acceptance band without exceptions.

For the urinary bladder, the showcased patient in figure 5.16 is the only case, where critical exceeding of the acceptance band occurs. Although the clinical constraint  $D_{max} < 100$  % is exceeded for four patients, patient GA2.6 is the only case where also the acceptance band is exceeded. The exceeding of the clinical constraint occurred only for patients, where the  $D_{max}$  of  $DD_{pCT}$  was very close to 2.7 Gy(RBE), i.e. 100 % of the prescribed dose per fraction, as shown in the figure.

For the OARs, the width of the acceptance band is mostly defined by the IQR of the shift uncertainty sampling, since the investigated OARs are not located around the distal ends of the proton beams. For the  $D_{15}$ , the acceptance bands thus also fully cover the errorbars of the BD-CTs. For the  $D_{max}$  for the urinary bladder, only BD-CT<sub>ST</sub> and BD-CT<sub>STB2</sub> in figure 5.16 are fully covered. The approximate widths of the acceptance bands are comparable in size to the widths of the acceptance bands for the HIT dataset. This can be seen for example for the figures 5.15 and 5.16. This hints towards comparable treatment plans in terms of intrinsic robustness, as the basic boundary conditions (i.e. the prescribed dose and dose constraints) are fulfilled by all plans.

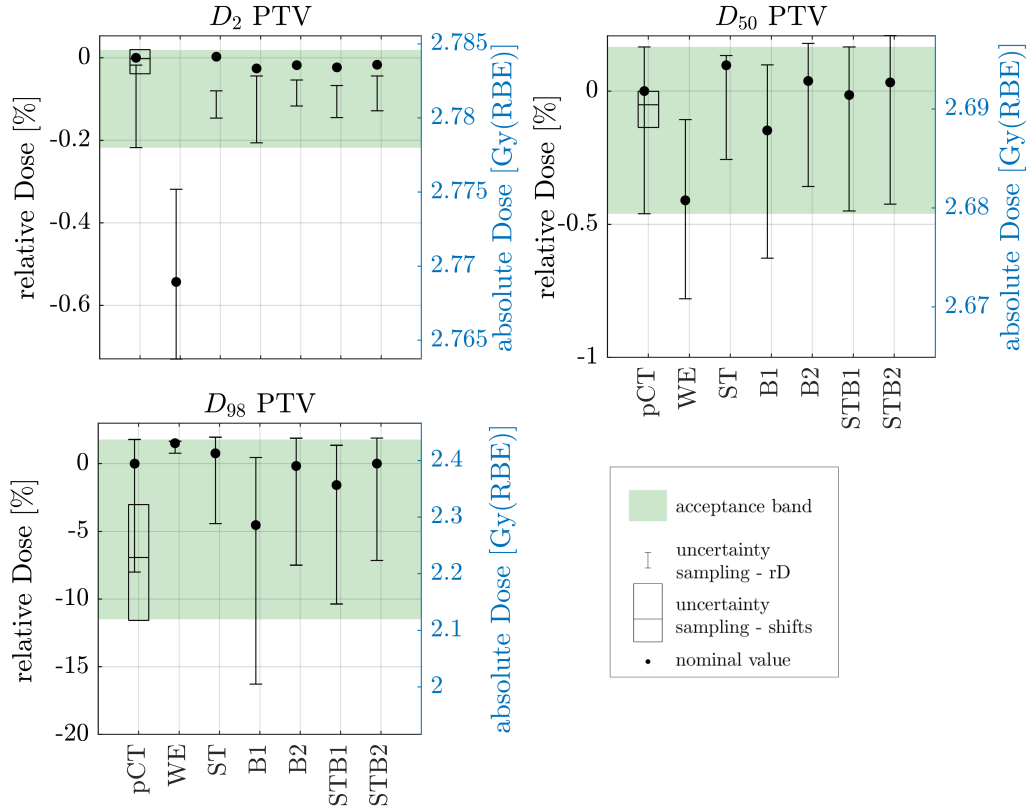


Figure 5.15.: DVMs for the pCT and BD-CTs for the PTV for patient GA2.6.

### 5.2.3. Discussion

**BD values** Generally, the comparison of Hounsfield unit (HU) values to literature HU values is more reasonable, as those are not biased by the applied HLUT. The analysis of the SPR values, on the other hand, allows comparison to methods that directly measure the SPR, for example proton radiographies (Schneider and Pedroni, 1995) and tomographies (Schulte et al., 2004). However, no such techniques are clinically available yet mostly due to technical reasons. Instead, only phantom measurements have been performed so far. Technically already possible is the usage of DECT images, from which SPR information can be derived (Wohlfahrt et al., 2017). SPR from DECT images is the closest to a ground truth achievable at the moment. DECT-derived SPR image would currently be the best reference for synthetic SPR images derived from MR. Unfortunately, no DECT image dataset or even a DECT with corresponding MR image dataset was available for this study. In anticipation of such data, the investigation in this thesis used SPR values instead of HU. The methods and the analysis procedure presented here are, therefore, suitable for future investigations, when MR and DECT data becomes available.

More confidence in the BD SPR values would in turn lead to a smaller uncertainty range from the rD uncertainty sampling. In this thesis,  $\pm 1.5\%$  was assumed as SPR

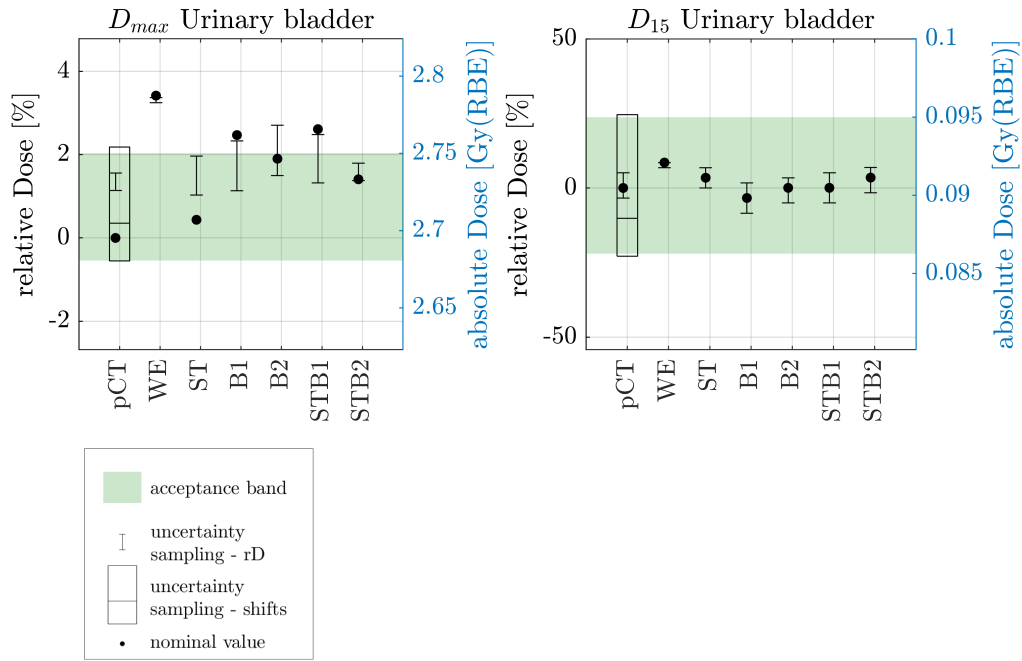


Figure 5.16.: DVMs for the pCT and BD-CTs for the urinary bladder for patient GA2.6.

or rD uncertainty, which is already a smaller uncertainty than presented in Paganetti (2012) for the combined uncertainties related to CT and HU-to-SPR conversion. Thus, with a smaller range of rD uncertainties, the remaining risk of exceeding the acceptance band would shrink to a minimum. The acceptance band, however, might shrink correspondingly where the rD uncertainty sampling has stronger impact than the set-up uncertainty sampling (e.g. for the PTV).

The comparison of the BD overwrite values derived from the SPR histogram shows some differences between the HIT and the Gold Atlas datasets: the fat BD value and the threshold are smaller for the HIT dataset than for the Gold Atlas, whereas for the three BD values for bone the opposite is the case. The most obvious explanation for this is the applied HLUT, that differed between the two datasets. It should, however, not be problematic for the investigation, as the BD values were derived from the SPR histogram instead of the HU histogram. This overcomes the bias of the applied HLUT relative to the pCT.

**BD-CT<sub>WE</sub>** The application of BD-CT<sub>WE</sub> is not recommended based on the poor performance of all the analysed metrics. Mean absolute range differences of more than 10 mm were observed for two HIT patients and nearly all Gold Atlas patients. The GPRs of the PTV drop down to almost 74 % and the DVM exceed the majority of acceptance bands. These findings are in agreement with Koivula et al. (2016), where the GPR of the WE sCT image was 86.8 % with a 2mm/2% criterion for the

volume where the dose is above 10 % of prescribed dose. It should be mentioned that this volume is larger than the PTV, which might explain the difference between the published GPR and the GPR observed here.

**Comparison of BD-CTs** The BD-CT<sub>ST</sub>, BD-CT<sub>B1</sub>, and BD-CT<sub>B2</sub> have been introduced to analyse the influence of different BD simplifications of the pCT individually. The BD-CT<sub>ST</sub> shows the worst results in all analysis metrics for the HIT dataset. Not surprisingly, that these faults propagate to the combined BD-CTs (BD-CT<sub>STB1</sub> and BD-CT<sub>STB2</sub>). For the Gold Atlas dataset, the BD-CT<sub>ST</sub> performed significantly better for all metrics, and, consequently, the BD-CT<sub>STB1</sub> and BD-CT<sub>STB2</sub> also perform better.

The source of faults of the BD-CT<sub>ST</sub> in the HIT dataset might be suboptimal choice of BD overwrite values. For example, the  $\mu_{ST}$  is influenced by organs that are not in the beam path, thus over-representing the corresponding SPR values and under-representing the SPR values of organs in the beam path. The soft tissue value should, theoretically, be more impaired by this, as most organs and soft tissue structures fall into this tissue class and are well-distinguishable from fat. However, when just comparing the SPR values, the fat SPR value seems more reasonable to be the source of dose differences than the soft tissue SPR value. The most reasonable explanation for the fault of the fat BD values might indeed be caused by the HLUT, but not the HLUT itself: if the peak of the fat HU values is Gaussian and a HLUT is applied, the peak might be skewed to one side. The peak position of the fitted Gaussian will then be slightly off.

The results obtained with the Gold Atlas dataset might be less sensitive by the specific choice of BD overwrite values, because the volumes of the PTV are much smaller and thus also the overall volume that the beams have to pass to reach the tumour. For comparison, the mean PTV volume of the HIT dataset is approximately 260 cm<sup>3</sup>, for the Gold Atlas dataset it is 67 cm<sup>3</sup>. The lack of a clinically contoured CTV as the basis for the PTV margin in the Gold Atlas dataset lead to this volume difference.

The most prominent problem of the BD-CT<sub>ST</sub> in the Gold Atlas dataset was the presence of a contrast agent. Its impact was mostly visible in the gamma cubes. At HIT, contrast agent is never applied for the acquisition of the pCT images, as iodine has a high atomic number and, thus, a high SPR. During the treatment, no contrast agent will be present and it would thus distort the planned DD significantly. Nyholm et al. (2018) state that the CT images in the Gold Atlas dataset were "acquired as part of clinical routine for treatment planning purposes". However, no information on the application of contrast agents and its purpose is given. This may explain the observed results for the BD-CT<sub>ST</sub> (without contrast agent) and the pCT (with contrast agent) for the Gold Atlas patients. Therefore, even better agreement between DD<sub>pCT</sub> and DD<sub>ST</sub> can be expected for pCT images without contrast agent.



**BD for urinary bladder** An exemplary improvement of the  $\text{BD-CT}_{\text{ST}}$  could be shown for patient HIT3 by adding a BD overwrite value for the urinary bladder. The source of the failure is the presence of urine in the urinary bladder, which is very similar to water. The threshold, however, classifies it as soft tissue, leading to a large difference in SPR in these voxels, which in turn leads to a large range difference whenever the beam passes the bladder. As the PTV volume in the Gold Atlas dataset is smaller, the beams hardly pass the urinary bladder. This explains why the negative effect of the urinary bladder seen for HIT patients was not observed in the 3D gamma cubes for the  $\text{DD}_{\text{ST}}$  for the Gold Atlas patients.

With regard to the MR images and the generation of sCT images from it, the bladder is not easily segmentable using grey value thresholds or probability masks as for bone. The reason is that the grey value in the MR is might not be unique for urine (depending on the sequence), and that the filling state can change the position and shape of the bladder dramatically. With an appropriate MR sequence or an advanced machine learning approach for sCT generation, a BD value for the bladder is feasible and, from the example shown here, also advisable, if the proton beam passes a considerable volume of the bladder.

**BDs for bone** The standard deviation of the BD values for bone are in the range of 0.010 and 0.026, which is quite large compared to the standard deviation of the soft tissue BD values. This indicates a large variability among the patients, which will become more important once cohort values will be applied. The  $\text{glsBD}$  values might be influenced by the segmentation method applied in this study. This was, however, consistently applied for all patients, thus the reason for the large standard deviation might be rather a biological variation in bone density between the patients. Larger cohorts need to be evaluated for more significant conclusions.

The results of the mean absolute range difference and the gamma analysis for  $\text{BD-CT}_{\text{B1}}$  and  $\text{BD-CT}_{\text{B2}}$  show no clear better option of the two, although the range difference plots indicate that  $\text{BD-CT}_{\text{B2}}$  might be the preferable solution. In the analysis of the DVMS, an improvement when going from  $\text{BD-CT}_{\text{B1}}$  to  $\text{BD-CT}_{\text{B2}}$  can be observed. It should be mentioned at this point, that a positive bias in the results of the images might be due to the usage of the pencil-beam algorithm (cf. section 2.3.2) for dose calculation. Vanderstraeten et al. (2007) performed a similar analysis investigating the number of tissue bins required for accurate Monte Carlo dose calculation. They recommend ten bins, i.e. ten BD values for bone. A further splitting of bone into smaller subsets with one BD each is feasible and can be expected to further improve the results also with the pencil-beam algorithm. However, it is more a diligent but routine piece of work to perform the BD-CT analysis for these BD-CTs. At this point, the interest rather increases in the question, how these BD values perform, when they are derived from an MR image.

**Comparison of the two datasets** Disregarding the differences for the BD-CT<sub>ST</sub>, the results obtained with the HIT dataset and the Gold Atlas dataset are very similar. The order of magnitude of the acceptance bands is comparable. However, the impact of the rD and the set-up uncertainty sampling differs between the datasets. For the  $D_{98}$  of the PTV, for example, the set-up sampling dominates the acceptance band, whereas for the Gold Atlas also the rD sampling has a significant impact. The  $D_{98}$  generally decreases for the Gold Atlas patients. Thus, the IQR lies below the nominal  $D_{98}$ . For the HIT dataset, the IQR spans around the nominal value. Two possible reasons might be the size of the PTV (where the large PTV makes it more robust against set-up errors) or the optimization itself being more robust to those errors.

The improved acceptance rate of the DVM values for the Gold Atlas might as well be due to the smaller volume. For both datasets, BD-CT<sub>STB2</sub> shows the better results of the two full-BD-CTs (BD-CT<sub>STB1</sub> and BD-CT<sub>STB2</sub>). The overall relative dose differences are below 1 % and only for a few cases, the acceptance band is exceeded. Thus, the BD content of BD-CT<sub>STB2</sub>, namely, air, fat, non-fatty soft tissue, soft bone and hard bone, will be the set of tissue classes, used to generate sCT images from MR images in the next step.

### 5.3. sCT from MR

This last section of the results summarizes the analysis results of the sCT images, that have been created as described in section 4.5.3 using the previously developed workflow for BD-CTs and the previously investigated set of BDs for air, fat, non-fatty soft tissue, soft bone, and hard bone. The results will also be compared to the BD-CT results, to identify the largest source of uncertainties and potential for improvement. The MR images from the Gold Atlas dataset have been modified to overcome the effect of the bias field (cf. section 2.1.2), they have been deformably registered to the CT images to analyse the impact of anatomical differences or distortions, and the two corrections of the MR images have been combined. In total, the original MR image and three corrected versions of the MR image (i.e. four in total) have been used to create four sCT images, respectively. As references, they are compared against the pCT and against the BD-CT<sub>STB2</sub>.

The basis for the sCT images were the T1-weighted MR images from the Gold Atlas dataset. Only the images from the 12 patients of site 1 and 3 were used, as the MR images are more comparable to each other in terms of the applied MR sequence. The metrics analysis for the DD<sub>STB2</sub> has been added to the results as well for those 12 patients only and might thus differ from the metric values presented in the section above, where all 19 patients were analysed.

#### 5.3.1. Range Analysis

The mean absolute range difference between the pCT and the BD-CT<sub>STB2</sub>, sCT<sub>O</sub>, sCT<sub>B</sub>, sCT<sub>R</sub>, and sCT<sub>BR</sub>, respectively, are summarized in table 5.11. It should be mentioned that the range analysis is independent from the structure delineations, as it only considers the pencil-beam positions planned for the PTV of the pCT.

Table 5.11.: Mean and standard deviation (sd) of the absolute difference in radiological depth between BD-CT<sub>STB2</sub>, the sCTs, and the pCT for all pencil-beams of the treatment plan for the 12 patients of the site 1 and site 3 of the Gold Atlas dataset. The average over all patients is shown in the bottom row.

patient	mean (sd) absolute range difference [mm]				
	STB2	O	B	R	BR
GA1.1-8	1.4 (1.0)	4.4 (2.9)	4.0 (2.8)	3.0 (2.1)	2.8 (2.1)
GA3.1-4	1.3 (0.9)	4.3 (2.5)	4.1 (2.5)	2.4 (2.0)	2.4 (2.0)
average	1.4 (1.0)	4.3 (2.7)	4.0 (2.7)	2.8 (2.1)	2.7 (2.0)

The difference between the two sites is small: a maximum deviation of 0.6 mm is observed for  $sCT_R$  between the two sites. For the deformably registered MR images, a significant improvement can be observed compared to the original MR image: from  $sCT_O$  to  $sCT_R$ , the absolute range difference decreases by 1.5 mm. However, the difference between  $sCT_R$  and  $BD-CT_{STB2}$  is still larger than 1 mm on average, indicating the impact of the conversion algorithm.

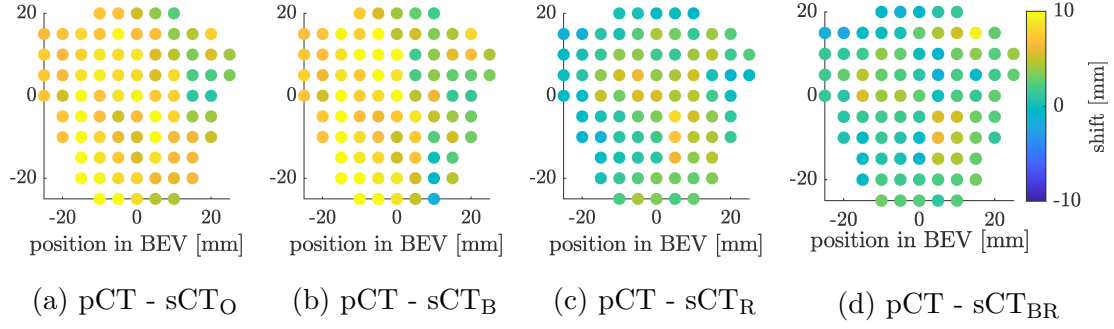


Figure 5.17.: Two-dimensional plots of the pencil-beam-wise range differences between sCTs and the pCT for patient GA3.3.

Figure 5.17 shows the 2D range difference plots for the same patient-case as for the BD-CT study in section 5.2.2. Comparing subfigures (a) and (b), or (c) and (d) to analyse the impact of the bias field correction (BFC), only small changes in the range difference can be seen. Subfigures (a) and (b) show large positive shifts all over the plot, which reduce significantly in subfigures (c) and (d). The remaining range differences for  $sCT_R$  and  $sCT_{BR}$  can be related to imperfect bone representation in the sCTs.

### 5.3.2. Gamma Analysis

The comparison of the  $DD_{pCT}$  with the  $DD_{STB2}$  and the  $DD_{sCTs}$  using the 3D gamma analysis is summarized in figure 5.18 for the PTVs. The gamma analysis is based on a voxel-to-voxel comparison between two DDs. For this reason, GPRs for delineated structures can only be evaluated for the exact same structures in both images. For the  $sCT_O$  and  $sCT_B$ , the structures differ from the structures in the pCT. Thus, only the structures from the pCT are considered during gamma analysis. The result will then represent changes in SPR in the path of the beam to that volume. Cases in which the PTV volumes moves out of the high dose region, for example due to a increase in bladder volume, however, will not be detected. The PTV in that case could be drastically underdosed, while the GPR could still be close to 100 %.

Although none of the sCTs reaches the same mean GPR as the  $BD-CT_{STB2}$ , the mean GPR for  $DD_R$  and  $DD_{BR}$  are above 97 %. The minimum GPRs for these two sCTs are 94.4 % and 93.4 %, respectively, for patient GA1.3.

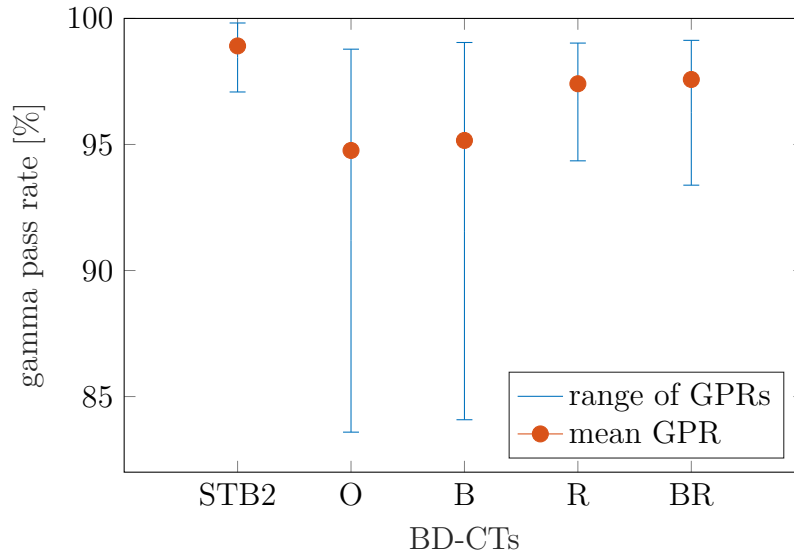


Figure 5.18.: Mean gamma pass rates (GPRs) using a 2mm/2% criterion for the PTV of  $DD_{STB2}$  and  $DD_{sCTs}$  (dots) for the Gold Atlas dataset. The ranges (minimum to maximum GPRs) are shown as errorbars.

Overall, the differences between the bias field corrected MR image and the respective uncorrected image are small, as previously also observed for the absolute range differences. The difference between the mean GPRs is 0.5 % for  $DD_O/DD_B$ , and 0.2 % for  $DD_R/DD_{BR}$ . Also the range of the GPRs seems almost unaffected by the BFC.

The mean GPR increases and the range of GPRs decreases, when registering the MR to the pCT. The mean GPR improves by approximately 3 percent points.

### 5.3.3. Dose Volume Histogram Analysis

The DVM differences from the  $DD_{sCTs}$  relative to the  $DD_{pCTs}$  averaged over all patients are shown in table 5.12 for the PTV, the urinary bladder, the rectum, and the RPW. The standard deviation of the 12 patients is given in brackets. For comparison, the results for  $DD_{STB2}$  are included.

Similar to the BD-CT results, the  $D_2$  and  $D_{50}$  are hardly impacted by the sCTs. For all images investigated, the (absolute) relative differences are below 0.5 %. For  $DD_{STB2}$ ,  $DD_R$  and  $DD_{BR}$  they are even below 0.1 %. For  $D_{98}$ , however, larger deviations can be observed up to approximately -40 % for individual patients.

The differences between  $DD_O$  and  $DD_B$ , and between  $DD_R$  and  $DD_{BR}$ , respectively, are small. This hints towards a minor impact of the BFC on the result. This finding is in agreement with previous findings for absolute range differences and GPRs.

The results for  $DD_O$  and  $DD_B$  clearly show larger deviations than  $DD_R$  and  $DD_{BR}$ .

Table 5.12.: Mean relative DVM difference and standard deviation (sd) of the 12 patients of the Gold Atlas dataset for the  $DD_{STB2}$  and the different  $DD_{sCTs}$ .

		mean (sd) relative DVM differences [%]									
	DVM	STB2		O		B		R		BR	
PTV	$D_2$	0.03	(0.07)	-0.22	(0.13)	-0.22	(0.10)	0.01	(0.10)	0.01	(0.11)
	$D_{50}$	0.02	(0.13)	-0.48	(0.29)	-0.50	(0.29)	0.08	(0.14)	0.05	(0.15)
	$D_{98}$	-1.55	(2.23)	-24.54	(13.22)	-24.64	(13.15)	-2.87	(4.60)	-3.55	(4.84)
urinary bladder	$D_{max}$	0.39	(0.73)	-2.60	(13.65)	-2.54	(13.53)	0.92	(0.87)	0.96	(0.99)
	$D_{15}$	-1.28	(1.50)	-34.73	(48.69)	-34.85	(48.56)	-0.92	(2.41)	-1.33	(2.08)
rectum	$D_{max}$	0.19	(0.64)	-9.55	(13.29)	-9.63	(13.26)	0.60	(0.58)	0.61	(0.58)
	$D_{15}$	-0.39	(0.69)	-24.42	(42.68)	-24.65	(42.77)	-0.95	(0.94)	-1.12	(0.84)
RPW	$D_{max}$	-0.02	(0.96)	-21.23	(34.31)	-21.12	(34.19)	0.75	(2.11)	0.63	(2.00)
	$D_{15}$	-0.70	(0.93)	-25.10	(37.53)	-25.10	(37.52)	-1.61	(1.69)	-1.59	(1.70)

Also the standard deviations are much larger. For some DVMs, the standard deviation already indicates that, although the majority of patients experiences a dose reduction, few cases also experience a dose increase relative to the nominal DD.

Figures 5.19 and 5.20 show an exemplary case for the DVMs of the PTV and the urinary bladder for patient GA1.3. This patient was chosen, as it was the only patient that showed a critical exceeding of the acceptance band for  $DD_R$  and  $DD_{BR}$ . It should be recalled that this patient also showed the lowest GPR for the PTV among all patients. Deviations as in figure 5.19 up to approximately -40 % for  $D_{98}$  occur for four patients. For 10 out of 12 patients,  $DD_O$  and  $DD_B$  fail the acceptance band criterion. The large range of rD uncertainty sampling for  $D_{98}$  for the sCT<sub>R</sub> and sCT<sub>BR</sub>, comparable to the respective range for BD-CT<sub>STB2</sub>, indicates a high risk of exceeding the acceptance band under just a slight uncertainty of rD.

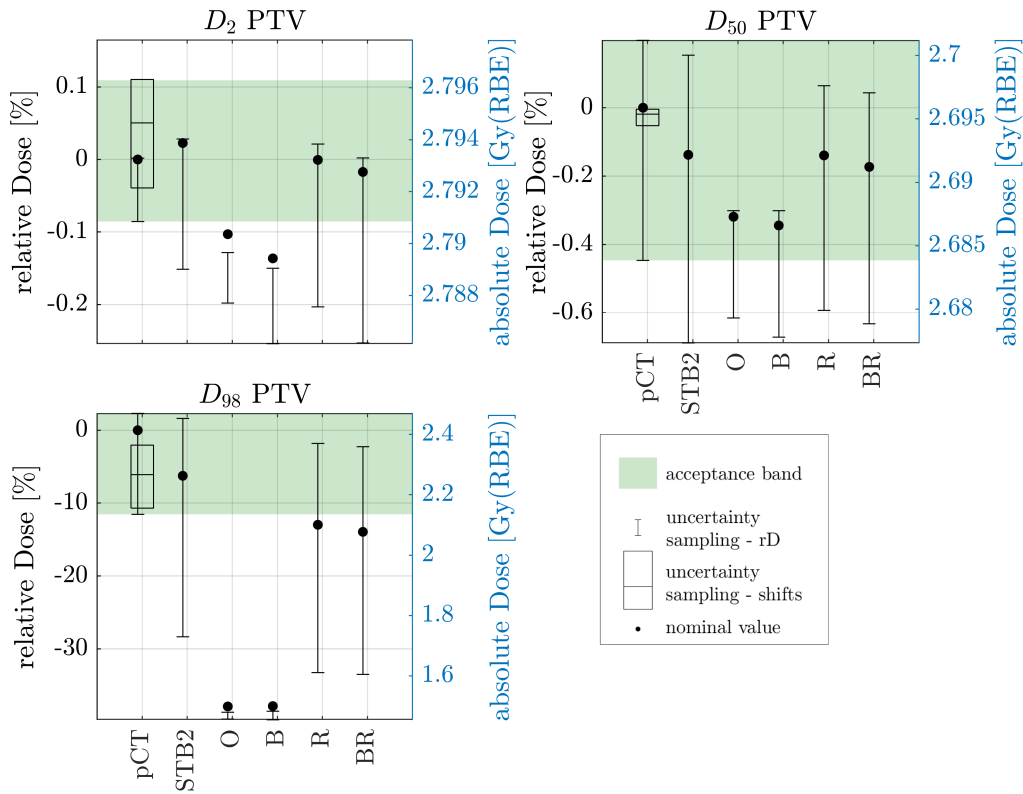


Figure 5.19.: DVMs for the pCT, BD-CT<sub>STB2</sub> and the sCTs for the PTV of patient GA1.3 for one fraction.

Also for the bladder significant DVM deviations of  $D_{max}$  and  $D_{15}$  (an increase in this case) can be observed for  $DD_O$  and  $DD_B$ . Thus, the showcased DVMs do not behave like the DVM parameter in table 5.12 averaged over the patient cohort. In this particular patient case, the bladder volume remained almost the same, however internal anatomical changes shift the target volume out of the high dose region, and the urinary bladder into the high dose region. For most patients in the Gold

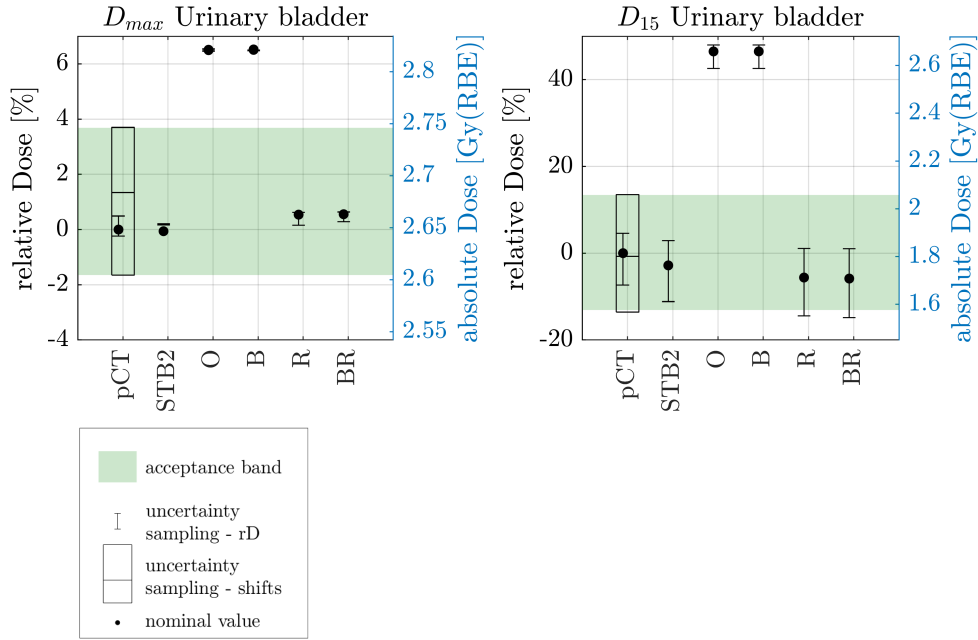


Figure 5.20.: DVMs for the pCT, BD-CT<sub>STB2</sub> and the sCTs for the urinary bladder of patient GA1.3 for one fraction.

Atlas dataset, the bladder volume increases, which also shifts the PTV out of the high dose region leading to a decrease in  $D_{98}$  and the urinary bladder into the high dose region. This would result in an increase of  $D_{max}$  for the urinary bladder and a decrease in  $D_{98}$  for the PTV. This can be observed for seven patients, for which the  $D_{max}$  values also exceed the acceptance bands. The mean relative difference for  $D_{max}$  reported in table 5.12, however, is around -2.5 %. This negative mean can be explained by the very strong decrease in  $D_{max}$  of around -30 % for patient GA1.4, which skews the mean to the negative.

The observable deviations of the DVMs for the DD<sub>R</sub> and DD<sub>BR</sub> are comparable to the BD-CTs. The relative differences of  $D_2$  and  $D_{50}$  are very small and not clinically relevant. For the  $D_{98}$ , larger deviations occur (as already shown in table 5.12), which are covered by the acceptance band for all patients except the showcased patient GA1.3. The errorbar for the rD uncertainty sampling is overall covered by the acceptance band, because the width of the acceptance band is dominated by the set-up uncertainty sampling.

The clinical constraint  $D_{max} < 100$  % of prescribed dose, i.e.  $< 2.7$  Gy(RBE) per fraction for the urinary bladder is exceeded by 10 patients, and by five patients for the rectum. Nonetheless, for no patient DD<sub>R</sub> or DD<sub>BR</sub> the DVMs exceed the acceptance bands. For DD<sub>O</sub> and DD<sub>B</sub>, seven, two, and one patient exceeded the acceptance bands for  $D_{max}$  for the urinary bladder, the rectum, and the RPW, respectively.



#### 5.3.4. Discussion

The presented sCT image generation is fast and robust because of the applied, simple algorithm. The analysis of the sCT images included the original MR and the corrected MR images. Two corrections were applied: a BFC to account for an MR-intrinsic artefact, and deformable image registration to account for anatomical changes in the patient between the CT and MR image acquisition. The latter is not a correction in the sense that the anatomical change is a false reflection of the true anatomy. However it impacts the comparison to the pCT as the ground truth. This correction yields a comparison between the pCT and the sCT which is only affected by the BD values and the conversion algorithm.

**Anatomical differences between MR and CT** Due to the few BD classes used for the sCT generation, only anatomical changes with respect to these classes can be detected in the comparison between  $DD_O$  and  $DD_R$ . Most prominently, body contour changes or shifts of bony structures are detected this way. Also interfraction changes such as weight loss, i.e. a reduction of fat, can be detected. Organ motion or changes such as a change in urinary bladder filling are hardly detectable without the respective BD class. This is confirmed by the large deviations observed for  $DD_O$  and  $DD_B$  for the DVMs, covering anatomical changes of individual structures. Similar observations can not be made in the range analysis and the gamma analysis, as those only considered pCT-related volumes. Situation, where for example the PTV for the GPR analysis shifts away from the high dose region, are not covered. In this case, the dose at the original PTV position is just influenced by the changes in SPR.

As the CT image generally is not applicable for structure delineation, it is advisable to use the underlying MR image at this point as well. Its better soft tissue contrast makes it more suitable for manual or even automated structure delineation. These delineations can then easily be propagated to the sCT image. Reasons for updated structures can be concerns with respect to the before-mentioned OARs, for example the urinary bladder or the rectum, moving into high dose regions.

The original MR showed anatomical deformations in comparison to the pCT image which could already be identified by visual inspections. The best parameter to quantify these deformations is the change of urinary bladder volume, as the bladder was segmented by clinicians on both the MR and the pCT. Among the 12 patients analysed, the highest MR to pCT bladder ratio was 2.1 for patient GA1.5, while the mean ratio for the dataset was 1.4. The results of the HIT dataset showed that an additional error in the DD occurs, when the beam passes through the urinary bladder, which was overwritten with a BD value for non-fatty soft tissue. This error also might occur for the  $sCT_O$  images for the Gold Atlas dataset, if the bladder volume increases such that it move towards the target volume. However, a reference pCT image with corresponding anatomy is missing. Thus, this error can not be

distinguished from the pure anatomical changes and adds to the uncertainty of the results. In future investigations, however, it is recommended to include the bladder as an individual BD class. For this purpose, adequate segmentation techniques for the MR image have to be developed.

To quantify the impact of the change in anatomy,  $sCT_R$ , which was deformably registered to the pCT was added to the analysis. The deformable registration process adds a source of uncertainty. The determination of the quality of the registration is difficult, especially if images of different modalities like CT and MR are compared. The large range of average magnitudes of residual displacements of fiducial markers after deformable registration (2.3 mm - 7.4 mm) for MR/CT image pairs of the prostate region presented by Brock (2010) indicate the difficulty of proper registration. The most intuitive method for determination of registration quality is the use of anatomical landmarks (Oh and Kim, 2017). Anatomical landmarks may include points defined by the observer in both images, whole organ delineations or implanted seeds. Since the pre-contoured bladder volume was already used for contour guidance during registration, it could not be utilized for analysis of registration quality. Nyholm et al. (2018) state that the registration parameter file in the Gold Atlas dataset was optimized based on visual inspection of the registered image. Thus, the uncertainty introduced by the registration can not be quantified for this specific study and the results using deformably registered MR images have to be handled with care.

For the mean absolute range difference and the GPR for the PTV, a clear improvement of results could be observed when applying the deformable image registration. The range difference decreased about 1.4 mm going from  $sCT_O$  to  $sCT_R$ . Also the standard deviation decreased, which indicates more reliable results. The GPR increased approximately by 3 percent points and the minimum GPR observed among the 12 patients increased from less than 80 % to more than 90 %. The shift uncertainty sampling of the pCT mostly covers the deviations of DVMs by  $sCT_R$  and  $sCT_{BR}$ , making the results acceptable with regard to the acceptance band. The relative DVM differences of  $sCT_O$  and  $sCT_B$  to the pCT show the drastic impact of anatomical changes as, for example, the increase in urinary bladder filling. Specifically, this change can shift the position of the prostate and thus the PTV by a few millimetres. As a result, the PTV is not in the high dose region any more and not covered adequately by the prescribed dose. This might indicate improper choice of the safety margin, i.e., the additional volume of the PTV compared to the CTV. At the same time, the urinary bladder moves into the high dose region, receiving higher total dose. It is however questionable whether such strong anatomical changes would actually be observed in the clinic, as most clinics have strict procedures regarding pre-treatment defaecation and urination. It would have been helpful to assess the observed differences, if the elapsed time between the CT and the MR acquisition had been reported, and if respective protocols have been followed also for the MR. Nonetheless, the observed deviations emphasize the need for pre-treatment imaging and measures to correct for them if necessary.

**Bias field correction** The second correction applied to the MR images, namely the BFC, showed minor impact of the quality on the sCT images. In other words, the number of BD classes is quite small. Thus, the identification of the classes is hardly affected by intra-class intensity variations. The small improvements observed for sCT<sub>B</sub> and sCT<sub>BR</sub> might be due to minor corrections, leading to a change of tissue class. However, the overall impact is still very small. The low impact of the BFC might be caused by the typical location of intensity deviations close to the radio-frequency (RF) coils at the anterior and posterior sides of the body (cf. section 2.1.2), while the beams typically enter the body laterally.

One clear advantage of the BFC, however, is its impact on the histogram of MR images. The example of a histogram shown in figure 4.7 is derived from a corrected MR image. For an uncorrected image, the two peaks are more smeared out, especially the one at higher intensity values. The valley position between the two peaks is then harder to distinguish. With respect to automated techniques, the valley position can easily be derived from the shown example. For this reason, BFC is generally advisable.

For sCT images with more BD classes or covering the full HU range, the impact of the bias field and its correction is worth investigating as well, as more grey value-based mis-classification can be expected.

**Impact of the different uncertainties** The differentiation of the uncertainties affecting the DD on a sCT image is a complex problem. Several aspects have to be considered: the use of BD values, the deformable image registration of the MR to the CT image (and the propagation of contours which are also applied for the pCT image), the approach for detecting the tissue classes in the MR image, and the anatomical changes within the body from the acquisition of the CT to the acquisition of the MR image.

While the use of BD values can be quantified accurately with the BD-CTs, the remaining uncertainties are harder to differentiate and quantify. The sCT<sub>R</sub> and sCT<sub>BR</sub> are affected by the use of BD values, the deformable image registration, and the detection of tissue classes in the MR. For example, a suboptimal choice of the separation threshold for the non-fatty soft tissue peak (lower intensity) and the fat peak (higher intensity) in the MR image can have systematic impact on the DD. A higher threshold leads to a misclassification of fat being overwritten with a BD for soft tissue. This increases the water-equivalent path length (WEPL) of the beam, leading to a shorter range. The same shorter range might occur, if due to the registration relatively more fat or less non-fatty soft tissue is in the beam path. Both are not distinguishable.

For the sCT<sub>O</sub> and sCT<sub>B</sub>, no registration error affects the results, however the change in anatomy can change the composition of tissue in the beam path, leading to different segmented tissue classes and different BD values in the beam path. The algorithm used for tissue class segmentation has the same impact (and uncertainties)

as for  $sCT_R$  and  $sCT_{BR}$ .

For all three investigated analysis metrics, namely the range difference analysis, the gamma analysis, and the analysis of relative DVM differences, the anatomical differences lead to a degradation in metric results as compared to the sCT from the rigidly registered MR image. The mean absolute range difference increases from approximately 2.7 mm to 4.0 mm, the GPR for the PTV drops from 97 % to 94 %, and all relative DVM difference increase in absolute terms going from  $sCT_R$  to  $sCT_O$ . This degradation of the DD could be avoided by the use of MR-image guidance of the treatment. To differentiate the impact quantitatively, either CT and MR have to be acquired directly after each other for minimal anatomical changes, or, less preferably, the registration error has to be quantified using applicable techniques.

The simulated range uncertainties showed that the sCT images would suffer from similar changes in DD as the pCT. The set-up uncertainty in particular should decrease significantly just by the application of the MR imaging and was therefore not sampled for the sCT images. As the range uncertainty had only minor impact on OARs, the risk of unintentional overdose due to uncertainties could clearly be reduced using the MR.

**Bone segmentation and BD values** From the visual inspection of the sCT image example shown in figure 4.9 and the comparison to the pCT and BD-CT images, it becomes clear that the identification of bone and its subclasses has potential for improvement. While the soft tissue classes show only minor differences in the images, large areas of the pubic bone (the central top bone) are not detected in the MR image. The threshold of 0.5 in the probability mask of bone applied in this study might introduce an uncertainty to the bone segment. In future studies, other (lower) thresholds should be tested and compared to the present study. Also the differentiation of soft bone and hard bone generates a rather spotty pattern, instead of solid hard bone areas. As mentioned before, previous studies indicated that more sub-classes within bone are necessary for accurate dose distributions (Vanderstraeten et al., 2007). The observed good performance with only two sub-classes in this study might originate from the applied dose-calculation engine. Studies comparing a pencil-beam algorithm with Monte Carlo techniques showed discrepancies in the DD particularly behind bones and at bone-tissue interfaces (Paganetti et al., 2008). Monte Carlo techniques could help to investigate the accuracy of the DDs in the bone and its sub-classes in sCT images in the future.

The probability mask for bone is certainly a useful tool. However, more patients should be included in its generation to make it more robust. It should also be noted that a gender-specific probability mask has to be generated, as the female pelvis and the pelvic bone have different anatomy than the male counterpart.

**Potential improvement of BD values** In the discussion in section 5.2.3 it was already mentioned that the usage of DECT images as pCT images would reduce the error of the HLUT. Generally, the comparison of both the BD-CTs and sCTs to the pCT might yield good results. However the difference to the actually delivered DD can not be estimated. Future investigations could possibly reduce this uncertainty. With the use of DECT images, or measurements of the SPR of the required tissue classes, e.g. fat, muscle, cortical bone, etc. For the latter, SPR could be measured directly via shoot-through measurements of monoenergetic proton beams. The measured values could then directly be applied as BD overwrites. Möhler et al. (2018) performed such experiments using pork tissue samples. Measurements of human tissue are still pending, and for various reasons can not be expected in the near future.

**Impact of the MR sequence** In the presented study, so far only a T1-weighted MR image was investigated. The advantage of this image is that such MR sequences are clinical standard and provide sufficient contrast between cortical bone, fat, and non-fatty soft tissue. Physicians are trained in interpreting these images and using them, for example, for organ delineations. Dedicated sequences such as ultra short echo time (UTE) sequences may improve bone segmentation. It should be investigated further, if the conversion of bone from MR to sCT improves when using a single UTE-MR-image as input for the sCT generation, without sacrificing the accuracy of the soft tissue classification. On the other hand, the absence of bone signal is also useful information to distinguish bone from other tissues. As bone is quite predictable in terms of location as compared to internal gas for example, this information can be used to make the separation between air and bone, as it was done in the presented study. The modular framework of this thesis would enable fast adaptation of the analysis workflow for comparison with the results presented in this study.

**Comparison to literature** Although several studies have been reported in the literature for sCT images for proton therapy, only two studies focused on prostate cancer (Koivula et al., 2016; Maspero et al., 2017). As the requirements for the image and the required granularity are different, for example, in the head region, direct comparison is only reasonable for the prostate studies. Koivula et al. (2016) applied the so-called dual model conversion technique. The dual model assigns a piece-wise linear conversion function for soft tissues, and a second-order polynomial conversion function for bone. They reported a mean GPR (2mm/2%) for the volume receiving more than 10 % of prescribed dose of 98.6 %. The minimum GPR of the 10 patients that were investigated was 97.2 %. Maspero et al. (2017) adapted a commercial BD overwrite technique with BD values for internal and outside air, soft tissue, fat, soft and hard bone. The BD values were derived from the literature. Compared to the commercial solution, they introduced the internal air class and adapted the

BD values for bones such that the range difference between the pCT and the sCT was minimized for two patient cases. They reported a mean GPR of 98.4 % and a minimum of 96.4 % with the same criteria as Koivula et al. (2016). In both studies, the mean GPR is about 1 percent point higher than in this study, also the minimum GPR is higher. It should however be noted, that the volume receiving more than 10 % of prescribed dose is much larger than the PTV volume investigated in this study. As the volume of GPR failure is fairly condensed for proton therapy around the target volume, an increase in GPR can be expected for a larger volume.

Maspero et al. (2017) also reported range analysis of the proton beams. However, including only average values of the median range differences of the individual patients. They reported a value of 0.1 mm, while no absolute range differences or minimum and maximum range differences for individual pencil-beams were reported. For this reason, a comparison to the presented values is not possible. The decrease in  $D_{98}$  which was observed here was less pronounced for both studies in comparison ( $< 1$  % estimated from the boxplots). Since Maspero et al. (2017) had very similar BD classes as in this study, the reasons for the stronger decrease observed here might most probably be either the bone segmentation technique or the choice of overwrite values. As further investigations have been recommended before in this study, bone is a promising candidate for improvement.

**Comments on the framework** The BFC correction were performed using the external software *Slicer*, the deformable image registration was performed using *elastix*. Although the deformable image registration is not required in a clinical workflow any more, both the BFC and the registration should be implemented in matlab for the integration into the matRad-based framework in the future. Combinations with existing tools, for example for deformable image registration in matlab, are conceivable.

## 6. Summary and Outlook

Magnetic resonance (MR)-guided proton therapy could reduce the uncertainties from set-up and inter- and intra-fractional anatomical changes, without an additional dose burden to the patient. MR images, however, lack stopping power ratio (SPR) information of the human tissue, which is necessary for dose calculation. To enable dose calculation, a possible solution is the generation of so-called synthetic computed tomography (sCT) images from the MR image.

The generalization and compatibility of sCT generation techniques is currently hindered by the fact that MR images from different clinics are not comparable. The reasons are varying MR scanners and/or MR sequences used at different clinics. From a practical point of view, MR images should ideally be acquired and processed when the patient is already in treatment position. Thus, the aim for this investigation was to generate sCT images for proton therapy of prostate cancer in a fast, robust and compatible fashion.

The required granularity of the tissue classification of the planning computed tomography (pCT) image was investigated in the first part of this study as a measure of robustness: the less tissue classes have to be identified in an MR image in the final step of this thesis, the lower is the probability of misclassification and the higher is the robustness of the approach. In the second part, a recommended set of tissue classes, i.e. the result of the granularity investigation in the first part, was identified in an MR image. The MR image was deformably registered to the pCT to overcome anatomical differences between the two images. A simple and fast conversion of MR to sCT image was applied: bones were separated from air via a grey value threshold in the MR histogram and a bone probability mask. The remaining tissue was segmented using only histogram-based thresholds. In the last part, the original MR image was converted to a sCT image using the same technique as in the second part.

For photon therapy, the granularity study showed that four tissue classes, namely air, fat, non-fatty soft tissue and bone, overwritten with bulk density (BD) values are dosimetrically indistinguishable from the pCT. Using photon therapy in the first place was used for framework development and as an independent verification of the methods and results, as much literature is available for comparison. For proton therapy, however, five tissue classes are required to yield good dosimetric agreement with the pCT, namely, air, fat, non-fatty soft tissue, soft bone and hard bone.

While grey-value based tissue classification for BD overwrite within the pCT is straightforward, the five tissue classes from the granularity study also had to be identified in the MR image. Especially in standard MR images, like the applied T1-weighted sequence, tissues with similar signal intensity like air and bone pose a challenge to distinguish. A solution based on a probability mask of bony tissue in the male pelvis accelerated and simplified the bone/air segmentation in the image. A fast and robust method for soft tissue identification was the application of a histogram-based threshold. The implementation of this approach was facilitated by the bias field correction (BFC), as the two main tissue classes fat and non-fatty soft tissue became better separable in the histogram. As the threshold is derived from the histogram dynamically, any sequence from any scanner can theoretically be used, as long as adequate water/fat separation is possible.

The developed modular framework is an excellent tool to test individual aspects of sCT generation. Upon demand, single modules can be exchanged and the overall results can be compared to previous results. This is particularly helpful, if, for example, two MR sequences or two methods to extract bone from the MR should be compared. Further applications are conceivable.

In this study, the comparison of the original, rigidly registered and the deformably registered MR images showed minor dose deviations of the sCT compared to the dosimetric impact of anatomical differences. The uncertainty sampling of the set-up and range uncertainties was introduced as a new method to evaluate and contextualize the observed dosimetric differences of the sCT images to the pCT. The large range of set-up uncertainties observed for some cases can be reduced by the application of MR-guidance for set-up and position verification, even during treatment.

Further improvement can be expected in future work when including a tissue class for the urinary bladder, and a finer subsampling for bone. A refinement of the bone segmentation in particular should be investigated in the future using dedicated sequences with a grey value representation for bony tissue, without neglecting the applicability of such sequences to segment soft tissues. Patient specific values for BDs can be derived from any pelvis computed tomography (CT) image available for the patient. Alternatively, the application of cohort values can be investigated in future studies. The BD values can also be derived prior to the MR image, not prolonging the procedure of sCT generation just before or during treatment.

In conclusion, sCT images could be successfully generated based on a clinically standard MR sequence and verified dosimetrically against the pCT. Few tissue classes, which have to be identified and replaced with BD values, could be derived robustly and fast from an MR image. MR-guidance for proton therapy is thus an applicable tool for treatment improvement. The technology for live imaging should be advanced in the near future to tap the full potential of MR-guidance.



# Appendices



# A. Deformable Image Registration

Parameter file for deformable image registrations using *elastix* (cf. section 4.3) used in this thesis to register MR images to CT images. The parameter file can be found in the data repository for the Gold Atlas dataset (more details can be found in the publication by Nyholm et al. (2018)).

```
(FixedInternalImagePixelType "float")
(MovingInternalImagePixelType "float")
(FixedImageDimension 3)
(MovingImageDimension 3)
(UseDirectionCosines "true")

// ***** Main Components *****
(Registration "MultiMetricMultiResolutionRegistration")
(ResampleInterpolator "FinalBSplineInterpolator")
(Resampler "OpenCLResampler")
(OpenCLResamplerUseOpenCL "true")
(FixedImagePyramid "FixedSmoothingImagePyramid"
"FixedSmoothingImagePyramid" )
(MovingImagePyramid "MovingSmoothingImagePyramid"
"MovingSmoothingImagePyramid" )
(Interpolator "BSplineInterpolator" "BSplineInterpolator" )
(Metric "NormalizedMutualInformation" "AdvancedMeanSquares")
(MetricOWeight 30000)
(Metric1Weight 1)
(Optimizer "AdaptiveStochasticGradientDescent")
(Transform "BSplineTransform")

// ***** Multiresolution *****
(NumberOfResolutions 3)
(ImagePyramidSchedule 4.0 4.0 4.0 2.0 2.0 2.0 1.0 1.0 1.0)

// ***** Transformation *****
(FinalGridSpacingInPhysicalUnits 60 60 60)
(GridSpacingSchedule 4 2 1)
(AutomaticScalesEstimation "true")
```

```

(AutomaticTransformInitialization "false")
(HowToCombineTransforms "Compose")

// ***** Optimizer *****
(MaximumNumberOfIterations 500 500 500)

// ***** Similarity measure *****
(NumberOfHistogramBins 64 96 96)
(ErodeMask "false")

// ***** Image sampling *****
(NumberOfSpatialSamples 3000)
(NewSamplesEveryIteration "true")
(ImageSampler "RandomCoordinate" "RandomCoordinate")
(CheckNumberOfSamples "true")
(RequiredRatioOfValidSamples 0.01)
(UseRandomSampleRegion "true")
(SampleRegionSize 150 150 70)
(MaximumNumberOfSamplingAttempts 100)

// ***** Interpolation and Resampling *****
(BSplineInterpolationOrder 2)
(FinalBSplineInterpolationOrder 3)
(DefaultPixelValue -1)
(WriteResultImage "false")
(ResultImagePixelType "short")
(ResultImageFormat "mhd")

```

The MR image was used as moving image, the CT image as fixed. The bladder contour in binary format (0/255) was used as second input image each, the binary body contour (0/255) derived from the CT images was used as mask. The registration command was

```

elastix -f0 <CT> -f1 <CT bladder> -fMask <CT contour> -m0 <MR>
-m1 <MR bladder> -out <result output directory> -p <parameter file>

```

The resulting transform file was then applied to the MR image and all binary segmentations of the MR

```

transformix -in <MR/segmentation> -out <transformed output directory>
-tp <result output directory\transform>

```

For the segmentation transformations, the `FinalBSplineInterpolationOrder` was changed from 3 to 0 to avoid artefacts when deforming binary images.

## B. List of Publications

### Conferences

Thesis findings have been presented in the following contributions to conferences:

Handrack, J., Bangert, M., Möhler. C., Bostel, T., Greilich, S. (2018). Simplification of relative electron density information in synthetic CT images for dose calculation. *ESTRO 37, Barcelona, Spain*. (Poster)

Handrack, J., Bangert, M., Möhler. C., Bostel, T., Greilich, S. (2018). Simplified CT images for dose calculation for pelvic tumour patients. *MR in RT, Utrecht, The Netherlands*. (Poster)

Handrack, J., Bangert, M., Möhler. C., Ellerbrock, M., Greilich, S. (2018). Synthetic CT accuracy for proton therapy of pelvic cancer. *DGMP Tagung, Nürnberg, Germany*. (Poster)

Handrack, J., Bangert, M., Möhler. C., Ellerbrock, M., Greilich, S. (2019). Assessment of bulk-density CT accuracy for MR-guided proton therapy of prostate cancer. *ESTRO 38, Milano, Italy*. (Poster)

### Manuscripts

In the following manuscript, findings of section [5.1](#) are presented:

Handrack, J., Bangert, M., Möhler. C., Bostel, T., Greilich, S. (2019). Towards a generalized development of synthetic CT images and assessment of their dosimetric accuracy. (in revision at *Acta Oncologica*)



# Bibliography

- Andreasen, D., Leemput, K. V., Hansen, R. H., Andersen, J. A. L., & Edmund, J. M. (2015). Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. *Medical Physics*, 42(4):1596–1605. [20](#)
- Austin-Seymour, M., Chen, G. T., Rosenman, J., Michalski, J., Lindsley, K., & Goitein, M. (1995). Tumor and target delineation: Current research and future challenges. *International Journal of Radiation Oncology Biology Physics*, 33(5):1041–1052. [19](#)
- Beavis, A. W., Gibbs, P., Dealey, R. A., & Whitton, V. J. (1998). Radiotherapy treatment planning of brain tumours using MRI alone. *The British Journal of Radiology*, 71(845):544–548. PMID: 9691900. [19](#)
- Berger, M., Hubbell, J., Seltzer, S., Chang, J., Coursey, J., Sukumar, R., Zucker, D., & Olsen, K. XCOM: Photon Cross Sections Database. Available at: <https://www.nist.gov/pml/xcom-photon-cross-sections-database> [Jul 2019]. [9](#)
- Bethe, H. (1930). Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie. *Annalen der Physik*, 397(3):325–400. [15](#)
- Bloch, F. (1946). Nuclear induction. *Physical Review*, 70:460–474. [4](#)
- Borras, J. M., Barton, M., Grau, C., Corral, J., Verhoeven, R., Lemmens, V., van Eycken, L., Henau, K., Primic-Zakelj, M., Strojan, P., Trojanowski, M., Dyzmann-Sroka, A., Kubiak, A., Gasparotto, C., Defourny, N., Malicki, J., Dunscombe, P., Coffey, M., & Lievens, Y. (2015). The impact of cancer incidence and stage on optimal utilization of radiotherapy: Methodology of a population based analysis by the ESTRO-HERO project. *Radiotherapy and Oncology*, 116(1):45–50. [1](#)
- Bortfeld, T., Schlegel, W., & Rhein, B. (1993). Decomposition of pencil beam kernels for fast dose calculations in three-dimensional treatment planning. *Medical Physics*, 20(2):311–318. [12](#)
- Bostel, T., Nicolay, N. H., Grossmann, J. G., Mohr, A., Delorme, S., Echner, G., Häring, P., Debus, J., & Sterzing, F. (2014). MR-guidance – a clinical study to evaluate a shuttle- based MR-linac connection to provide MR-guided radiotherapy. *Radiation Oncology*, 9(1). [25](#)

- Bostel, T., Pfaffenberger, A., Delorme, S., Dreher, C., Echner, G., Haering, P., Lang, C., Splinter, M., Laun, F., Müller, M., Jäkel, O., Debus, J., Huber, P. E., Sterzing, F., & Nicolay, N. H. (2018). Prospective feasibility analysis of a novel off-line approach for MR-guided radiotherapy. *Strahlentherapie und Onkologie*, 194(5):425–434. [25](#)
- Brock, K. K. (2010). Results of a Multi-Institution Deformable Registration Accuracy Study (MIDRAS). *International Journal of Radiation Oncology Biology Physics*, 76(2):583 – 596. [88](#)
- Bushberg, J. T., Seibert, J. A., Jr., E. M. L., & Boone, J. M. (2012). *The essential physics of medical imaging*, (3. ed.). Wolters Kluwer, Lippincott Williams and Wilkins. [10](#)
- Chen, L., Price, R. A., Wang, L., Li, J., Qin, L., McNeeley, S., Ma, C. M. C., Freedman, G. M., & Pollack, A. (2004). MRI-based treatment planning for radiotherapy: Dosimetric verification for prostate IMRT. *International Journal of Radiation Oncology Biology Physics*, 60(2):636–647. [19](#), [20](#)
- Chen, S., Qin, A., Zhou, D., & Yan, D. (2018). Technical Note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. *Medical Physics*, 45(12):5659–5665. [21](#)
- Demol, B., Boydev, C., Korhonen, J., & Reynaert, N. (2016). Dosimetric characterization of MRI-only treatment planning for brain tumors in atlas-based pseudo-CT images generated from standard T1-weighted MR images. *Medical Physics*, 43(12):6557–6568. [20](#)
- Diwakar, M. & Kumar, M. (2018). A review on CT image noise and its denoising. *Biomedical Signal Processing and Control*, 42:73–88. [34](#)
- Doemer, A., Chetty, I. J., Glide-Hurst, C., Nurushev, T., Hearshen, D., Pantelic, M., Traughber, M., Kim, J., Levin, K., Elshaikh, M. A., Walker, E., & Movsas, B. (2015). Evaluating organ delineation, dose calculation and daily localization in an open-MRI simulation workflow for prostate cancer patients. *Radiation Oncology*, 10(1):37–37. [37](#)
- Edmund, J. M., Kjer, H. M., Leemput, K. V., Hansen, R. H., Andersen, J. A. L., & Andreasen, D. (2014). A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times. *Physics in Medicine and Biology*, 59(23):7501–7519. [2](#), [22](#)
- Edmund, J. M. & Nyholm, T. (2017). A review of substitute CT generation for MRI-only radiation therapy. *Radiation Oncology*, 12(1). [2](#), [33](#)



- Eilertsen, K., Vestad, L. N. T. A., Geier, O., & Skretting, A. (2008). A simulation of MRI based dose calculations on the basis of radiotherapy planning CT images. *Acta Oncologica*, 47(7):1294–1302. [20](#), [57](#)
- Fiorino, C., Reni, M., Bolognesi, A., Bonini, A., Cattaneo, G. M., & Calandrino, R. (1998a). Set-up error in supine-positioned patients immobilized with two different modalities during conformal radiotherapy of prostate cancer. *Radiotherapy and Oncology*, 49(2):133 – 141. [18](#)
- Fiorino, C., Reni, M., Bolognesi, A., Cattaneo, G. M., & Calandrino, R. (1998b). Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiotherapy and Oncology*, 47(3):285–292. [17](#)
- Han, X. (2017). Mr-based synthetic ct generation using a deep convolutional neural network method. *Medical Physics*, 44(4). [21](#)
- Hong, L., Goitein, M., Bucciolini, M., Comiskey, R., Gottschalk, B., Rosenthal, S., Serago, C., & Urie, M. (1996). A pencil beam algorithm for proton dose calculations. *Physics in Medicine and Biology*, 41(8):1305–1330. [16](#)
- Hsu, S.-H., Cao, Y., Huang, K., Feng, M., & Balter, J. M. (2013). Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiation therapy. *Physics in Medicine and Biology*, 58(23):8419–8435. [21](#)
- ICRU (2010). Prescribing, recording, and reporting photon-beam IMRT, Report 83. *Journal of the International Commission on Radiation Units and Measurements*, 10(1). [19](#)
- Johansson, A., Karlsson, M., & Nyholm, T. (2011). Ct substitute derived from mri sequences with ultrashort echo time. *Medical Physics*, 38(5):2708–2714. [21](#)
- Johnstone, E., Wyatt, J. J., Henry, A. M., Short, S. C., Sebag-Montefiore, D., Kelly, C. G., McCallum, H. M., & Speight, R. (2017). A systematic review of synthetic CT generation methodologies for use in MRI-only radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 100(1):199–217. [2](#)
- Jonsson, J. H., Karlsson, M. G., Karlsson, M., & Nyholm, T. (2010). Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiation Oncology*, 5(1):62–62. [20](#), [37](#)
- Kapanen, M. & Tenhunen, M. (2012). T1/T2\*-weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning. *Acta Oncologica*, 52(3):612–618. [21](#)

- Khoo, V., Dearnaley, D., Finnigan, D., Padhani, A., Tanner, S., & Leach, M. (1997). Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning. *Radiotherapy and Oncology*, 42:1–15. [19](#)
- Klein, S., Staring, M., Murphy, K., Viergever, M. A., & Pluim, J. P. (2010). elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205. [29](#), [32](#)
- Koivula, L., Wee, L., & Korhonen, J. (2016). Feasibility of MRI-only treatment planning for proton therapy in brain and prostate cancers: Dose calculation accuracy in substitute CT images. *Medical Physics*, 43(8):4634–4642. [2](#), [21](#), [22](#), [77](#), [91](#), [92](#)
- Korsholm, M. E., Waring, L. W., & Edmund, J. M. (2014). A criterion for the reliable use of MRI-only radiotherapy. *Radiation Oncology*, 9(16). [20](#), [37](#)
- Kraus, K. M., Jäkel, O., Niebuhr, N. I., & Pfaffenberger, A. (2017). Generation of synthetic ct data using patient specific daily mr image data and image registration. *Physics in Medicine and Biology*, 62(4):1358–1377. [20](#)
- Lee, Y. K., Bollet, M., Charles-Edwards, G., Flower, M. A., Leach, M. O., McNair, H., Moore, E., Rowbottom, C., & Webb, S. (2003). Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. *Radiotherapy and Oncology*, 66(2):203–216. [19](#), [20](#), [37](#)
- Lomax, A. J. (2008). Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. *Physics in Medicine and Biology*, 53(4):1043–1056. [1](#), [17](#)
- Low, D. A., Harms, W. B., Mutic, S., & Purdy, J. A. (1998). A technique for the quantitative evaluation of dose distributions. *Medical Physics*, 25(5):656–661. [42](#)
- Marks, L. B., Yorke, E. D., Jackson, A., Haken, R. K. T., Constine, L. S., Eisbruch, A., Bentzen, S. M., Nam, J., & Deasy, J. O. (2010). Use of normal tissue complication probability models in the clinic. *International Journal of Radiation Oncology Biology Physics*, 76(3):S10–S19. [11](#), [53](#)
- Maspero, M., Savenije, M. H. F., Dinkla, A. M., Seevinck, P. R., Intven, M. P. W., Jurgenliemk-Schulz, I. M., Kerkmeijer, L. G. W., & van den Berg, C. A. T. (2018). Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Physics in Medicine and Biology*, 63(18):185001. [21](#)
- Maspero, M., van den Berg, C. A. T., Landry, G., Belka, C., Parodi, K., Seevinck, P. R., Raaymakers, B. W., & Kurz, C. (2017). Feasibility of MR-only proton dose calculations for prostate cancer radiotherapy using a commercial pseudo-CT generation method. *Physics in Medicine and Biology*, 62(24):9159–9176. [2](#), [22](#), [91](#), [92](#)

- Müller, L.-R. (2016). Commissioning of an Uncertainty Analysis Tool for IMPT. Master's thesis, Department of Physics and Astronomy, University of Heidelberg. [46](#)
- Möhler, C., Russ, T., Wohlfahrt, P., Elter, A., Runz, A., Richter, C., & Greulich, S. (2018). Experimental verification of stopping-power prediction from single- and dual-energy computed tomography in biological tissues. *Physics in Medicine and Biology*, 63(2):025001. [91](#)
- Nyholm, T., Nyberg, M., Karlsson, M. G., & Karlsson, M. (2009). Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments. *Radiation Oncology*, 4(1):54–54. [17](#), [19](#)
- Nyholm, T., Svensson, S., Andersson, S., Jonsson, J., Sohlén, M., Gustafsson, C., Kjellén, E., Söderström, K., Albertsson, P., Blomqvist, L., Zackrisson, B., Olsson, L. E., & Gunnlaugsson, A. (2018). MR and CT data with multiobserver delineations of organs in the pelvic area-Part of the Gold Atlas project. *Medical Physics*, 45(3):1295–1300. [23](#), [26](#), [29](#), [30](#), [32](#), [78](#), [88](#), [97](#)
- Oh, S. & Kim, S. (2017). Deformable image registration in radiation therapy. *Radiation Oncology Journal*, 35(2):101–111. [31](#), [88](#)
- Paganetti, H. (2012). Range uncertainties in proton therapy and the role of monte carlo simulations. *Physics in Medicine and Biology*, 57(11):R99–R117. [77](#)
- Paganetti, H., Jiang, H., Parodi, K., Slopsema, R., & Engelsman, M. (2008). Clinical implementation of full Monte Carlo dose calculation in proton beam therapy. *Physics in Medicine and Biology*, 53(17):4825–4853. [90](#)
- Pileggi, G., Speier, C., Sharp, G. C., Garcia, D. I., Catana, C., Pursley, J., Amato, F., Seco, J., & Spadea, M. F. (2018). Proton range shift analysis on brain pseudo-CT generated from T1 and T2 MR. *Acta Oncologica*, 57(11):1521–1531. [2](#), [22](#)
- Rank, C. M., Hünemohr, N., Nagel, A. M., Röthke, M. C., Jäkel, O., & Greulich, S. (2013). MRI-based simulation of treatment plans for ion radiotherapy in the brain region. *Radiotherapy and Oncology*, 109(3):414–418. [2](#), [21](#), [22](#)
- Roberson, P. L., McLaughlin, P. W., Narayana, V., Troyer, S., Hixson, G. V., & Kessler, M. L. (2005). Use and uncertainties of mutual information for computed tomography/magnetic resonance (CT/MR) registration post permanent implant of the prostate. *Medical Physics*, 32(2):473–482. [17](#)
- Roeske, J. C., Forman, J. D., Mesina, C. F., He, T., Pelizzari, C. A., Fontenla, E., Vijayakumar, S., & Chen, G. T. (1995). Evaluation of changes in the size and location of the prostate, seminal vesicles, bladder, and rectum during a course of external beam radiation therapy. *International Journal of Radiation Oncology Biology Physics*, 33(5):1321–1329. [18](#)

- Schad, L. R., Blüml, S., Hawighorst, H., Wenz, F., & Lorenz, W. J. (1994). Radio-surgical treatment planning of brain metastases based on a fast, three-dimensional MR imaging technique. *Magnetic Resonance Imaging*, 12(5):811–819. [20](#)
- Schaffner, B. & Pedroni, E. (1998). The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between CT-HU and proton stopping power. *Physics in Medicine and Biology*, 43(6):1579–1592. [17](#)
- W. Schlegel, C. Karger, & O. Jäkel, ed. (2018). *Medizinische Physik*. Springer Spektrum. [14](#)
- Schneider, U. & Pedroni, E. (1995). Proton radiography as a tool for quality control in proton therapy. *Medical Physics*, 22(4):353–363. [76](#)
- Schulte, R., Bashkirov, V., Tianfang Li, Zhengrong Liang, Mueller, K., Heimann, J., Johnson, L. R., Keeney, B., Sadrozinski, H. F. ., Seiden, A., Williams, D. C., Lan Zhang, Zhang Li, Peggs, S., Satogata, T., & Woody, C. (2004). Conceptual design of a proton computed tomography system for applications in proton radiation therapy. *IEEE Transactions on Nuclear Science*, 51(3):866–872. [76](#)
- Siversson, C., Nordström, F., Nilsson, T., Nyholm, T., Jonsson, J., Gunnlaugsson, A., & Olsson, L. E. (2015). Technical Note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm. *Medical Physics*, 42(10):6090–6097. [2](#), [20](#)
- Spadea, M. F., Pileggi, G., Zaffino, P., Salome, P., Catana, C., Izquierdo-Garcia, D., Amato, F., & Seco, J. (2019). Deep Convolution Neural Network (DCNN) multi-plane approach to synthetic CT generation from MR images - application in brain proton therapy. *International Journal of Radiation Oncology Biology Physics*. [2](#), [21](#), [22](#)
- Tanner, S. F., Finnigan, D. J., Khoo, V. S., Mayles, P., Dearnaley, D. P., & Leach, M. O. (2000). Radiotherapy planning of the pelvis using distortion corrected MR images: the removal of system distortions. *Physics in Medicine and Biology*, 45(8):2117–2132. [19](#)
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320. [32](#)
- Tyagi, N., Fontenla, S., Zhang, J., Cloutier, M., Kadbi, M., Mechalakos, J., Zelefsky, M., Deasy, J., & Hunt, M. (2017). Dosimetric and workflow evaluation of first commercial synthetic CT software for clinical use in pelvis. *Physics in Medicine and Biology*, 62(8):2961–2975. [2](#), [20](#), [57](#), [58](#)

- Uh, J., Merchant, T. E., Li, Y., Li, X., & Hua, C. (2014). MRI-based treatment planning with pseudo CT generated through atlas registration. *Medical Physics*, 41(5). 20
- van Herk, M. (2004). Errors and margins in radiotherapy. *Seminars in Radiation Oncology*, 14(1):52–64. 46
- Vanderstraeten, B., Chin, P. W., Fix, M., Leal, A., Mora, G., Reynaert, N., Seco, J., Soukup, M., Spezi, E., Neve, W. D., & Thierens, H. (2007). Conversion of CT numbers into tissue parameters for monte carlo dose calculations: a multi-centre study. *Physics in Medicine and Biology*, 52(3):539–562. 79, 90
- Weishaupt, D., Köchli, V. D., Marincek, B., Froehlich, J. M., Nanz, D., & Prueßmann, K. P. (2014). *Wie funktioniert MRI?*, (7., überarb. und erg. Aufl. ed.). Springer. Engl. Ausg.: How does MRI work? 6
- White, D. R., Griffith, R. V., & Wilson, I. J. (2016). Report 46. *Journal of the International Commission on Radiation Units and Measurements*, os24(1). 37, 56
- Wieser, H.-P., Cisternas, E., Wahl, N., Ulrich, S., Stadler, A., Mescher, H., Müller, L.-R., Klinge, T., Gabrys, H., Burigo, L., Mairani, A., Ecker, S., Ackermann, B., Ellerbrock, M., Parodi, K., Jäkel, O., & Bangert, M. (2017). Development of the open-source dose calculation and optimization toolkit matRad. *Medical Physics*, 44(6):2556–2568. 29
- Wiesinger, F., Bylund, M., Yang, J., Kaushik, S., Shanbhag, D., Ahn, S., Jonsen, J. H., Lundman, J. A., Hope, T., Nyholm, T., Larson, P., & Cozzini, C. (2018). Zero TE-based pseudo-CT image conversion in the head and its application in PET/MR attenuation correction and MR-guided radiation therapy planning. *Magnetic Resonance in Medicine*. 21
- Wilson, R. R. (1946). Radiological use of fast protons. *Radiology*, 47(5). 1
- Witte, M. G., van der Geer, J., Schneider, C., Lebesque, J. V., Alber, M., & van Herk, M. (2007). IMRT optimization including random and systematic geometric errors based on the expectation of TCP and NTCP. *Medical Physics*, 34(9):3544–3555. 46
- Wohlfahrt, P., Möhler, C., Stützer, K., Greilich, S., & Richter, C. (2017). Dual-energy CT based proton range prediction in head and pelvic tumor patients. *Radiotherapy and Oncology*, 125(3):526–533. 76
- Woodard, H. Q. & White, D. R. (1986). The composition of body tissues. *The British Journal of Radiology*, 59(708):1209–1218. 17
- Wu, J., Haycocks, T., Alasti, H., Ottewell, G., Middlemiss, N., Abdoell, M., Warde, P., Toi, A., & Catton, C. (2001). Positioning errors and prostate motion during

conformal prostate radiotherapy using on-line isocentre set-up verification and implanted prostate markers. *Radiotherapy and Oncology*, 61(2):127 – 133. [18](#)

Zheng, W., Kim, J. P., Kadbi, M., Movsas, B., Chetty, I. J., & Glide-Hurst, C. K. (2015). Magnetic Resonance–Based Automatic Air Segmentation for Generation of Synthetic Computed Tomography Scans in the Head Region. *International Journal of Radiation Oncology Biology Physics*, 93(3):497–506. [21](#), [22](#)

## **Erklärung**

Ich erkläre hiermit, dass ich die vorgelegte Dissertation selbst verfasst und mich dabei keiner anderen, als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Heidelberg, den 16.09.2019

.....